

Kognitive Fähigkeiten und Mathematikleistungen
im Grundschulalter –
Kreuzverzögerte Effekte über vier Messzeitpunkte

Dissertation zur Erlangung des akademischen Grades
Doctor philosophiae (Dr. phil.)
der Philosophischen Fakultät
der Universität Rostock
im Fach Psychologie

Dissertationsschrift

eingereicht am 26.04.2017

von

Daniela Hoese, geb. Fiedler

Matrikel-Nr.: 5200991

Gutachter

1. Gutachter: Prof. Dr. Christoph Perleth | Institut für Pädagogische Psychologie
„Rosa und David Katz“ | Philosophische Fakultät | Universität
Rostock
2. Gutachter: Prof. Dr. Andrea Hildebrandt | Institut für Psychologie | Fakultät für
Psychologie | Differentielle und Persönlichkeitspsychologie/
Psychologische Diagnostik | Universität Greifswald
3. Gutachter: PD Dr. Dorothee Doerfel-Baasen | Institut für Pädagogische
Psychologie „Rosa und David Katz“ | Philosophische Fakultät |
Universität Rostock

Datum der Einreichung: 26. April 2017

Datum der Verteidigung: 19. Juli 2017

Danksagung

Zuallererst möchte ich mich bei all den Menschen bedanken, von denen das Gelingen dieser Dissertation abhängig war und die mich während der letzten Jahre auf so vielfältige Weise wesentlich unterstützt und begleitet haben.

Prof. Dr. Christoph Perleth danke ich für die fachliche Betreuung, die konstruktiven Hinweise von Beginn an und seine, stets von Geduld geprägte, umfassende Betreuung während der gesamten Promotionszeit. Er hat mir nicht nur die Möglichkeit gegeben, diese Arbeit anzufertigen, sondern mich auch von vielen anderweitigen Aufgaben entlastet. Seine stets wertvollen und konstruktiven Rückmeldungen haben diese Arbeit zu einer besseren gemacht. Auch für die Hilfe zur fachlichen Qualifizierung, die Unterstützung auf Tagungen, die wissenschaftlichen Anregungen und hilfreichen Hinweise möchte ich ihm meinen herzlichen Dank aussprechen.

Prof. Dr. Andrea Hildebrandt danke ich außerordentlich für die umfassende und fundierte Betreuung, die wissenschaftlichen Denkanstöße und die präzisen, detaillierten und äußerst hilfreichen Rückmeldungen von der Fragestellung bis zur Fertigstellung dieser Arbeit. Insbesondere die Hinweise zur qualifizierten Datenanalyse und Interpretation trugen wesentlich dazu bei, dass diese Arbeit in ihrer jetzigen Form vorgelegt werden kann. Zudem möchte ich mich bei ihr für die wertvolle persönliche und fachliche Begleitung im Rahmen des Doktorandinnen-Mentoring Programms bedanken. Sie hat mir als Betreuerin und Ansprechpartnerin stets ihre einfühlsame und geduldige Unterstützung angeboten.

PD Dr. Dorothee Doerfel-Baasen danke ich sehr für ihre persönliche Unterstützung, den fachlichen Rat und das Interesse an meiner Arbeit sowie die Bereitschaft, diese zu begutachten.

Prof. Dr. Ana Altaras Dimitrijević möchte ich für das Korrekturlesen, das Redigieren der Texte und die exzellenten Rückmeldungen ganz besonders danken.

Allen Kollegen des Instituts für Pädagogische Psychologie "Rosa und David Katz" gilt nicht nur mein besonderer Dank für die fachliche Unterstützung, sondern auch dafür, dass sie mir persönlich und emotional stets zur Seite gestanden haben. Prof. Dr. Eva Stumpf, Dr. Rosina Neumann, Steffen Winny, Angelika Haase, Ulrike Bruhn, Nadine Becker-Hingst, Fabian Jobst, PD Dr. Dorothee Doerfel-Baasen und Prof. Dr. Ana Altaras Dimitrijević möchte ich sagen: Es ist für mich von unschätzbarem Wert, mit Ihnen und Euch zusammenarbeiten zu können! An dieser Stelle möchte ich auch den ehemaligen Mitarbeiterinnen des Instituts meinen herzlichen Dank aussprechen.

Frau Marit Schmidt möchte ich ganz besonders herzlich danken. Sie hat dieses Vorhaben in allen Phasen mitgetragen, mir mit ihrer liebevollen Art stets kompetent und persönlich zur Seite gestanden und mich daran erinnert, dass es wichtig ist, die persönliche Mitte zu behalten.

Ein besonderer Dank geht an Frau Prof. Dr. Katja Koch und Herrn Prof. Dr. Bodo Hartke und ihren Mitarbeiterinnen und Mitarbeitern, die im Rahmen der Mecklenburger Längsschnittstudie tätig waren und dadurch die Umsetzung dieses Dissertationsvorhabens überhaupt erst ermöglicht haben.

Rostock, 26. April 2017

Abstract

In dieser Arbeit stehen die für die Forschungs- und pädagogische Praxis relevante Entwicklung kognitiver Leistungen und deren Zusammenhänge zu schulischen Leistungen im Grundschulalter im Vordergrund.

Dabei sind theoretische Annahmen zur zeitlichen Präzedenz zwischen kognitiven Fähigkeiten und mathematischen Leistungen kontrovers. Vorliegende Theorien reichen von der Annahme einer dominierenden Beeinflussung von Mathematikleistungen durch Intelligenz, über gegenseitige Beeinflussung, im Sinne des Matthäus-Effekts (Stanovich, 1986), bis hin zu der Betrachtung von Schulleistungen und kognitiven Leistungen als überlappende Konstrukte (Übersicht bei Watkins et al., 2007). Bislang berichtete Längsschnittdaten zu diesem Thema beschränken sich zumeist auf zwei Messzeitpunkte und lassen keine Aussage über den zeitlichen Verlauf kreuzverzögerter Effekte zu.

Werden kognitive Fähigkeiten und mathematische Fähigkeiten differenziert betrachtet, ist bislang ungeklärt, inwiefern beide Konstrukte womöglich identische Fähigkeiten erfassen. Über die faktorielle Differenziertheit der Intelligenz und schulischer Leistungen wurden erhebliche Debatten geführt (siehe Flanagan & McGrew, 1997). Diese Debatten gehen von drei Möglichkeiten aus: Vorstellungen, dass es sich bei Mathematikleistung und Intelligenz um überlappende Konstrukte handelt, werden bei Ceci (1991), innerhalb Deutschlands auch von Rindermann (2006), in der Gegenargumentation wiederum von Baumert, Brunner, Lüdtke & Trautwein (2007) heftig diskutiert. Eine Übersicht geben Watkins et al. (2007). Die vorliegende Arbeit versucht daher mit der ersten Fragestellung einen grundlegenden theoretischen Beitrag zur Klärung bisheriger kontroverser theoretischer Annahmen zu leisten: Erfolgt die Erfassung kognitiver Fähigkeiten und der Mathematikleistungen innerhalb der Messzeitpunkte partiell unabhängig? Inwiefern handelt es sich folglich bei Mathematikleistungen und kognitiven Fähigkeiten um unabhängige Konstrukte?

Zuverlässige Aussagen lassen sich vorzugsweise aus Längsschnittdaten ableiten, aber nur wenige Längsschnittstudien berücksichtigen situationsspezifische bzw. messfehlerbedingte Einflüsse, deren unerwünschte Kumulationen über einen längeren Untersuchungszeitraum zu verzerrten Schlussfolgerungen führen können. Die Überprüfung, inwiefern kognitive Fähigkeiten und Mathematikleistungen über einen Zeitraum von vier Messzeitpunkten im jährlichen Abstand vergleichbar gemessen werden können, ist insofern relevant, als dass die starke Veränderlichkeit der kindlichen kognitiven Entwicklung (Hof-

stätter, 1954; Schneider & Hasselhorn, 2007) Adaptionen der Testverfahren verlangt. Solche testverfahrensbedingten Varianzen in Messwerten bergen jedoch die Gefahr der verzerrten Interpretation der tatsächlichen Veränderung. Daher klärt die zweite Fragestellung zunächst, inwieweit die Messung kognitiver Fähigkeiten und der Mathematikleistungen über die Zeit vergleichbar ist.

Das Interesse gilt jedoch auch der Stabilität des Niveaus kognitiver Leistungen, welches ab dem frühen Erwachsenenalter eine hohe Niveaustabilität erreicht (Ziegler, 2004; Rost, 2010; Schneider & Hasselhorn, 2007). Für das Grundschulalter scheint die Struktur kognitiver Leistungen ab dem 10. Lebensjahr annähernd stabil zu sein (Rost, 2010; Trost, 1998; Franić, 2014), wobei der schulischen und familiären Lernumwelt sowie den nicht kognitiven Persönlichkeitsmerkmalen (z. B. Perleth, 1997) eine große Bedeutung hinsichtlich der weiteren Steigerung, insbesondere hinsichtlich der Positionsveränderungen im Vergleich mit Gleichaltrigen beigemessen wird. Im Zentrum der dritten Fragestellung dieser Arbeit soll daher beantwortet werden, inwieweit die erfassten Konstrukte kognitive Fähigkeiten und mathematische Leistungen im Grundschulalter als stabil zu betrachten sind.

Etwas außerhalb der bisherigen Forschungsbemühungen stand die Frage nach der gegenseitigen Beeinflussung von Schulleistungsmerkmalen und kognitiven Fähigkeiten. Der Verlauf kreuzverzögerter Effekte von kognitiven Fähigkeiten auf spätere Mathematikleistungen unter gleichzeitiger Berücksichtigung der gegenteiligen Effekte (von früheren Mathematikleistungen auf kognitive Fähigkeiten) ist noch nicht hinreichend untersucht worden und so werden theoretische Annahmen über die zeitliche Präzedenz zwischen den genannten Konstrukten immer noch kontrovers diskutiert. Von einer Beeinflussung der Mathematikleistungen vornehmlich durch Intelligenz gehen hingegen Lubinski und Dawis (1992) aus. Aber auch Annahmen einer gegenseitigen Beeinflussung von Mathematikleistungen und kognitiven Fähigkeiten im Kindes- und Jugendalter sind ebenso formuliert worden (Stanovich, 1986; Schneider & Stefanek, 2004). Bislang berichtete Längsschnittuntersuchungen umfassen nur zwei Messzeitpunkte und lassen nur ungenaue Aussagen zu dem zeitlichen Verlauf kreuzverzögerter Effekte zu. Die vierte Fragestellung wird daher wie folgt formuliert: Sind vor allem kognitive Fähigkeiten prädiktiv für Mathematikleistungen oder kann von einer gegenseitigen Beeinflussung ausgegangen werden?

Im jährlichen Abstand zu vier Messzeitpunkten wurden kognitive Fähigkeitstests der KFT-Testverfahrenreihe von Abou-Koura und Perleth (2005), Perleth und Heller (2008) und Heller und Perleth (2000), einerseits und mathematische Schulleistungen andererseits

(Deutscher Mathematiktest für erste, zweite, dritte und vierte Klassen der DEMAT-Testverfahrenreihe von Krajewski, Küspert & Schneider (2002), Krajewski, Liehm & Schneider (2004) sowie Roick, Gölitze & Hasselhorn (2004, 2006)) erhoben. Eine regionale Totalerhebung (2006-2009) mit einer Beteiligung von 1726 Grundschulkindern der ersten bis vierten Klassenstufen bildete die Datengrundlage.

Für die Klärung der konzeptionellen ersten Fragestellung wurden Zwei-Faktoren-Korrelationsmodelle auf latenter Ebene zu den jeweiligen zeitgleich eingesetzten Messverfahren sowohl auf der Gesamtkonstrukt- als auch auf der Subtestebene spezifiziert. Für die Invarianzüberprüfung (Fragestellung 2) wurden konstruktsspezifische Latent-State Modelle spezifiziert. Die Überprüfung der zeitlichen Stabilität (Fragestellung 3) erfolgte mit konstruktsspezifischen autoregressiven Strukturgleichungsmodellen, die anschließend in ein gemeinsames latentes Variablenmodell mit kreuzverzögerten Effekten (Fragestellung 4) zusammengeführt wurden. Dieses Modell sollte Aufschluss über die gegenseitige Beeinflussung kognitiver Leistungen und Mathematikleistungen geben.

Hinsichtlich der Prüfung der faktoriellen Differenziertheit (Fragestellung 1) der Konstrukte kognitive Fähigkeiten und mathematische Schulleistungen lässt sich im Ergebnis feststellen, dass dieselben Konstrukte miteinander über die vier Messzeitpunkte im jährlichen Abstand untereinander jeweils höher korrelieren als zum anderen Konstrukt des jeweiligen Zeitpunktes. Die latenten Variablen der anderen Konstrukte korrelieren dagegen im mittleren Bereich miteinander. Somit kann eine partielle Unabhängigkeit beider Konstrukte festgestellt werden.

Die Messung der Konstrukte kognitive Fähigkeiten und mathematische Leistungen erwies sich über die Grundschulzeit hinweg als invariant (Fragestellung 2), wobei die Invarianz für die kognitiven Fähigkeitstests im Vergleich zu den Testverfahren zur Erfassung der Mathematikleistungen konsistenter ist.

Die kognitiven Fähigkeiten wiesen eine hohe Stabilität (Fragestellung 3) der individuellen Merkmalsausprägungen auf. Die autoregressiven Pfade der ersten, zweiten und dritten Ordnung zusammen mit den Residualvarianzen zeigen Stabilitätseffekte zwischen .76 und .86. Diese Resultate zeigen, dass kognitive Fähigkeiten nicht nur kognitive Fähigkeiten des Folgejahres vorherzusagen vermögen, sondern sich auch über zwei bzw. drei Jahre (hier Messzeitpunkte im jährlichen Abstand) prognostizieren lassen. Für die mathematischen Fähigkeiten können ebenso Folgejahre durch die Messung hoher Stabilitätseffekte anhand der autoregressiven Pfade zwischen den mathematischen Fähigkeiten über

die Zeit vorhergesagt werden. Dennoch zeigen beide Konstrukte statistisch bedeutsame Residualvarianzen über die Zeit, die eine Überprüfung des Verlaufs der gegenseitigen Beeinflussung beider Konstrukte zulassen. Die kreuzverzögerten Effekte (Fragestellung 4) zeigen eine zunehmende Vorhersagekraft über die Jahre. Diese Ergebnisse unterstützen die theoretische Ansicht, dass im Grundschulalter keine einseitige Beeinflussung durch kognitive Fähigkeiten auf schulische Mathematikleistungen, sondern eine wechselseitige Beeinflussung auch von Mathematikleistungen auf kognitive Fähigkeiten empirisch zu beobachten ist.

Mit diesem empirischen Nachweis der präzisen Trennbarkeit der Konstrukte kann diese Arbeit einen theoretischen Beitrag zu den bislang heftig diskutierten Theoriekontroversen leisten. Testverfahren, denen ein hierarchisches, bereichsspezifisches Theorieverständnis zugrunde liegt sowie fortschrittliche Analysemethoden leisten einen entscheidenden Beitrag.

Um präzise Ergebnisse zu gewährleisten, sollten im Rahmen von Längsschnittstudien zusätzliche Qualitäten Berücksichtigung finden, die in Querschnittuntersuchungen nicht unbedingt erforderlich sind. Insbesondere im Kindesalter sind Testadaptionen für die entsprechende Alterskohorte eine Notwendigkeit, die allerdings bei Nichtberücksichtigung der Fehlerwerte in Längsschnittuntersuchungen die Gefahr von Fehlinterpretationen bergen. Dieses Risiko steigt mit der Anzahl der in der Studie eingesetzten Adaptionen. Die Konstruktkonsistenz der über die Zeit eingesetzten Testverfahren sollte daher bereits vor dem Einsatz der Messinstrumente berücksichtigt werden, sofern längsschnittliche Fragestellungen im Grundschulalter im Mittelpunkt stehen sollen. Auch bei der Bewertung vormaliger Längsschnittdaten zu diesem Thema ist die Berücksichtigung der Messfehler qualitätsbestimmend. Um zuverlässige Aussagen aus Längsschnittdaten abzuleiten, sollten Längsschnittstudien situationsspezifische bzw. messfehlerbedingte Einflüsse auf latenter Ebene berücksichtigen, deren unerwünschte Kumulationen bei Unterlassung der Kontrolle über einen längeren Untersuchungszeitraum zu verzerrten Schlussfolgerungen führen können.

Auf der Ebene der Konstrukte kann über den betrachteten Vierjahreszeitraum sowohl von einer hohen Stabilität der kognitiven Fähigkeiten als auch von einer hohen Stabilität mathematischer Leistungen im Grundschulalter ausgegangen werden. Aussagen zur interindividuellen Stabilität können zuverlässig formuliert werden, wenn Längsschnittdaten mit vier Messzeitpunkten aus unabhängigen Stichproben herangezogen werden können und die Analysen auch auf latenter Ebene durchgeführt werden. Die nachweisbaren kreuzverzö-

gerten Effekte lassen den begründeten Schluss zu, dass nicht nur kognitive Fähigkeiten spätere Mathematikleistungen beeinflussen, sondern dass auch vorwissensbasierte und wissensbasierte mathematische Leistungen einen deutlichen Einfluss auf spätere kognitive Fähigkeiten besitzen, deren Einfluss im Laufe der Grundschulzeit sogar zunimmt. Folgestudien sollten daher den Einfluss wissensbasierter Fähigkeiten einbeziehen. Latent-State-Spezifikationen erlauben dahingehend zuverlässige Analysen.

Inhaltsverzeichnis

Danksagung

Abstract

1	Einleitung	14
2	Theoretischer Teil.....	20
2.1	Kognitive Fähigkeiten und mathematische Schulleistungen im Grundschulalter.....	20
2.1.1	Kognitive Fähigkeiten	21
2.1.1.1	Forschungstraditionen in der Strukturanalyse von kognitiver Fähigkeiten	24
2.1.1.2	Testverfahren zur Messung kognitiver Fähigkeiten	46
2.1.2	Mathematische Schulleistungen	54
2.1.2.1	Entwicklung mathematischer Fähigkeiten.....	58
2.1.2.2	Mathematische Fähigkeiten in kognitiven Strukturmodellen.....	63
2.1.2.3	Prädiktoren von Mathematikleistungen.....	67
2.1.2.4	Messung und Bewertung von schulischen Leistungen.....	71
2.1.2.5	Schulleistungstests.....	72
2.1.3	Zwischenresümee kognitive Fähigkeiten und mathematische Schulleistungen.....	73
2.2	Ableitung der Fragestellungen.....	75
2.2.1	Fragestellung 1 – Faktorielle Differenziertheit	77
2.2.1.1	Forschungsposition – Identität	78
2.2.1.2	Forschungsposition – Faktorielle Differenziertheit.....	84
2.2.1.3	Faktorielle Differenziertheit – Zusammenfassung, Ableitung und Hypothesen	89
2.2.2	Fragestellung 2 – Invarianz	97
2.2.2.1	Invarianz – Zusammenfassung, Ableitung und Hypothesen	103
2.2.3	Fragestellung 3 – Autoregressive Effekte kognitiver Fähigkeiten und mathematischer Schulleistungen	104
2.2.3.1	Autoregressive Effekte der kognitiven Entwicklung über die Zeit	105
2.2.3.2	Autoregressive Effekte mathematischer Schulleistungen über die Zeit.....	113

2.2.3.3	Autoregressive Effekte – Zusammenfassung, Ableitung und Hypothesen	115
2.2.4	Fragestellung 4 – Kreuzverzögerte Effekte mathematischer Schulleistungen und kognitiver Fähigkeiten	118
2.2.4.1	Forschungsdebatten zu kreuzverzögerten Effekten kognitiver Fähigkeiten und schulischer Leistungen.....	141
2.2.4.2	Kreuzverzögerte Effekte – Zusammenfassung, Ableitung und Hypothesen	147
3	Empirischer Teil	153
3.1	Stichprobe	156
3.2	Studiendesign	162
3.3	Messinstrumente.....	163
3.3.1	Kognitive Fähigkeitstestverfahren der KFT – Reihe.....	163
3.3.1.1	KFT 1-2 R.....	164
3.3.1.2	KFT 3 R.....	166
3.3.1.3	KFT 4-12 R.....	168
3.3.2	Mathematische Schulleistungstests der DEMAT-Reihe	169
3.4	Analysemethoden	172
3.4.1	Analysestrategie 1 – Faktorielle Differenziertheit.....	175
3.4.2	Analysestrategie 2 – Invarianzprüfung.....	178
3.4.3	Analysestrategie 3 – Autoregressive Effekte.....	184
3.4.4	Analysestrategie 4 – Kreuzverzögerte Effekte	186
4	Ergebnisse	190
4.1	Ergebnisse 1 – Faktorielle Differenziertheit.....	205
4.1.1	Korrelationen manifeste Faktoren auf Subtestebene: KFT 1, KFT 2 – DEMAT 1, DEMAT 2.....	205
4.1.2	Korrelationen manifester Faktoren auf Subtestebene KFT 3, KFT 4 – DEMAT 3, DEMAT 4	207
4.1.3	Faktorielle Differenziertheit (latent) – KFT 1 vs. DEMAT 1	208
4.1.4	Faktorielle Differenziertheit (latent) – KFT 2 vs. DEMAT 2	209
4.1.5	Faktorielle Differenziertheit (latent) – KFT 3 vs. DEMAT 3	210
4.1.6	Faktorielle Differenziertheit (latent) – KFT 4 vs. DEMAT 4	211

4.1.7 Faktorielle Differenziertheit – Zusammenfassung und Diskussion der Ergebnisse.....	214
4.1.8 Zusammenfassung der Ergebnisse vor dem Hintergrund der Forschungsfragen	216
4.2 Ergebnisse 2 – Invarianz – KFT	225
4.2.1 Konfigurale Invarianz – KFT	225
4.2.2 Metrische Invarianz – KFT	230
4.2.3 Skalare Invarianz – KFT.....	231
4.2.4 Invarianz Modellvergleich – KFT	232
4.3 Ergebnisse 2 – Invarianz DEMAT	233
4.3.1 Konfigurale Invarianz – DEMAT	233
4.3.2 Metrische Invarianz – DEMAT	237
4.3.3 Skalare Invarianz – DEMAT	239
4.3.4 Invarianz Modellvergleich – DEMAT	239
4.3.5 Invarianz – Zusammenfassung und Diskussion der Ergebnisse	242
4.4 Ergebnisse 3 – Autoregressive Effekte.....	244
4.4.1 Autoregressive Effekte – 1. Ordnung.....	244
4.4.2 Autoregressive Effekte – 2. Ordnung.....	245
4.4.3 Autoregressive Effekte – 3. Ordnung.....	246
4.4.3.1 Autoregressive Effekte – Passungswerte.....	248
4.4.3.2 Autoregressive Effekte – Kontrolle der Mehrebenenstruktur	248
4.4.4 Autoregressive Effekte – Zusammenfassung und Diskussion der Ergebnisse.....	249
4.5 Ergebnisse 4 – Kreuzverzögerte Effekte	250
4.5.1 Kreuzverzögerte Effekte – Einflüsse kognitiver Fähigkeiten auf spätere mathematische Schulleistungen.....	250
4.5.2 Kreuzverzögerte Effekte – Hinzunahme der Einflüsse mathematischer Schulleistungen auf kognitive Fähigkeiten	252
4.5.3 Kreuzverzögerte Effekte – unter Hinzuziehung der komplexen Beziehungsmuster.....	254
4.5.4 Kreuzverzögerte Effekte – Zusammenfassung und Diskussion der Ergebnisse.....	264
5 Gesamtdiskussion, Implikationen und Perspektiven	267

6	Verzeichnisse	274
6.1	Literaturverzeichnis	274
6.2	Abbildungsverzeichnis	295
6.3	Tabellenverzeichnis	297
7	Anhang	302
7.1	KFT Invarianzprüfung	302
7.2	Invarianzprüfung DEMAT	311
7.3	Publikationsliste	320
7.4	Inhalt der beiliegenden CD	322

1 Einleitung

Schulische Leistungserbringung gilt als bewertbarer Maßstab für den Grad an schulisch erworbenem Wissen und nimmt eine gesellschaftlich dominierende Rolle ein, da der Schulerfolg eine wesentliche Voraussetzung für die Teilhabe am gesellschaftlichen Leben darstellt. Mathematische Schulleistungen gelten in dieser Hinsicht als bedeutendster Prädiktor für allgemeinen Schulerfolg und nehmen daher aufgrund ihrer pädagogischen Relevanz eine besondere Stellung ein. Kognitive Fähigkeiten gelten als Determinanten von Lern- und Wissenserwerbsprozessen (Watkins, Lei & Canivez, 2007), aber wie und in welchem Maße sie Schulleistungen beeinflussen, ist noch nicht endgültig geklärt. Bisherige Forschungsergebnisse geben keine eindeutige Antwort. Uneinigkeit herrscht bereits bei der Frage, ob es sich bei kognitiven Fähigkeiten und mathematischen Schulleistungen um identische, überlappende oder unabhängig voneinander existierende Konstrukte handelt.

Darüber hinaus gehen die Meinungen, wie stabil sich einmal erworbenes Wissen über die Grundschulzeit erweist und wie abhängig es seinerseits von den kognitiven Fähigkeiten ist, auseinander. Denn obwohl auch kognitive Fähigkeiten im Grundschulalter als relativ stabil gelten, liegen dafür bisher keine eindeutigen empirischen Belege vor. Unklar bleibt außerdem, inwiefern die kognitiven Fähigkeiten wiederum von früher erworbenem Wissen beeinflusst werden. Bisher vorliegende Studien, die zur Beantwortung herangezogen werden könnten, weisen häufig methodische Unstimmigkeiten auf und können keine belastbaren Ergebnisse liefern.

Für die pädagogische Praxis ist die Beantwortung der Fragen allerdings zentral, denn die Antwort kann eingrenzen, inwiefern vorhandenes Lernpotenzial optimal gesteigert werden kann und wovon diese Steigerung im Einzelnen abhängig ist. Mit diesen Informationen lässt sich dann entscheiden, welche Interventionen sinnvoll und möglich sind.

Die hier vorgestellte Arbeit möchte daher einen Beitrag zur Beantwortung der Fragen leisten und damit die theoretische Debatte bereichern. Sie setzt sich zunächst grundlegend mit den relevanten Theorien zu den betreffenden Konstrukten auseinander, um dann bisher vorliegende Untersuchungen zum Zusammenhang zwischen kognitiven Fähigkeiten und Schulleistungen kritisch zu diskutieren.

Das zentrale Ziel der vorliegenden Arbeit ist es, im Anschluss die noch ungeklärten Zusammenhänge zwischen kognitiven Fähigkeiten und mathematischen Schulleistungen in den Mittelpunkt des Forschungsinteresses zu stellen und im empirischen Teil belastbare

Ergebnisse über die komplexen Zusammenhänge zwischen kognitiven Fähigkeiten und mathematischen Schulleistungen über einen im Längsschnitt betrachteten Vierjahreszeitraum für die Grundschule aufzuzeigen. Im engen Zusammenhang damit wird die Zuverlässigkeit der über die Zeit verwendeten Messverfahren selbst hinterfragt. Häufig handelt es sich nämlich um Konzepte, die primär für Quer- und nicht Längsschnittanalysen konzipiert worden sind. In ihnen bleibt der Einfluss von Messfehlern über die Zeit oft unberücksichtigt, sodass kumulierte Messfehler Gefahr laufen, als reale Veränderungen interpretiert zu werden. Die vorliegende Arbeit ist somit nicht die erste, die sich mit den Zusammenhängen zwischen kognitiven Fähigkeiten und mathematischen Schulleistungen auseinandersetzt. Sie ist jedoch eine der ersten, die anstrebt, die Eignung von Testverfahren, die im Rahmen von Längsschnittanalysen eingesetzt werden, hinsichtlich ihrer Konstruktkonsistenz zu prüfen.

Für die Beantwortung der genannten zentralen Fragestellungen gliedert sich die Arbeit in einen theoretischen und einen empirischen Teil.

Im theoretischen Teil wird Bezug auf Theorien und Forschungsbefunde zu den Konstrukten kognitive Fähigkeiten und (mathematische) Schulleistungen genommen. Dabei erfolgt eine Konzentration auf das Grundschulalter, in dem sich Zusammenhänge zwischen kognitiven Fähigkeiten und schulischen Leistungen als besonders relevant erweisen. Häufig werden beispielsweise in diesem Zeitraum Interventionen angestrebt oder es stehen schulische Laufbahnentscheidungen an.

Die Eingrenzung auf mathematische Schulleistungen im empirischen Teil der Arbeit soll der Tatsache Rechnung tragen, dass insbesondere diese Beziehungen im Zentrum von Forschungsbemühungen stehen, da sie als relevantester Prädiktor späteren Schulerfolgs gelten (Ackerman, 2003; McGrew, Flanagan, Keith & Vanderwood, 1997; Übersicht bei Watkins et al., 2007). Da kognitive Fähigkeiten gleichermaßen mit schulischem Erfolg assoziiert werden (Heller, 1991; Renkl & Stern, 1994; Gottfredson, 1998), entstehen allein aus diesem Fakt die kontroversen Debatten um die Identität, die Unabhängigkeit bzw. die Überlappung der Konstrukte kognitive Fähigkeiten und mathematische Schulleistungen.

Zunächst wird im theoretischen Teil der Arbeit im Kapitel 2 das Problemfeld kognitive Fähigkeiten auf der einen Seite und mathematische Schulleistungen auf der anderen Seite abgesteckt. Im Kapitel 2.1.1 liegt der Fokus auf den kognitiven Fähigkeiten. Es werden die Begriffe kognitive Fähigkeiten und Intelligenz definiert und eine Auswahl der wichtigsten Modelle vorgestellt. Dabei erfolgt eine Eingrenzung auf Intelligenzstrukturmodelle, die geeignet sind, kognitive Fähigkeiten und mathematische Schulleistungen abzubilden.

Anschließend wird auf relevante Theorien und Befunde zum Stand der Forschung hinsichtlich kognitiver Fähigkeiten eingegangen. Explizit sollen Modelle, die Wissenskomponenten bzw. Vorwissen und Wissens Elemente berücksichtigen, Erwähnung finden, da ihr Einfluss über die Zeit zunehmend als wesentlich erachtet wird (Ericsson, 1991; Perleth, 1997; Helmke & Weinert, 1997). Zudem sollen Lernprozesse, die im Laufe der Zeit wirksam werden können, im Sinne eines dynamischen Verständnisses in die Argumentation einfließen. Anschließend sollen Testverfahren zur Messung kognitiver Fähigkeiten betrachtet werden sowie der Blick auf die Konzeptualisierung von Intelligenztestverfahren gerichtet werden. In Kapitel 2.1.2 folgt im Anschluss die Erörterung der Theorien und Befunde zu mathematischen Schulleistungen im Grundschulalter. Zunächst werden mathematische Schulleistungen als Teil von Schulleistungen begrifflich definiert sowie Faktoren mathematischer Schulleistungen abgebildet. Anschließend erfolgt ein Blick auf den Forschungsstand, der den Fokus auf die Zusammenhänge der Konstrukte kognitive Fähigkeiten und (mathematische) Schulleistungen legt.

Nach einem Zwischenresümee, in dem die wichtigsten eruierten Befunde zu den kognitiven Fähigkeiten sowie den mathematischen Schulleistungen kompakt zusammengefasst werden, werden im Hinblick auf das Hauptanliegen der Arbeit in Kapitel 2.2 vier zentrale Forschungsfragen formuliert.

Die erste Fragestellung stellt die Messungen kognitiver Fähigkeiten und mathematischer Schulleistungen in den Mittelpunkt und damit die Frage, ob es sich um identische, einander überlappende oder voneinander unabhängige Konstrukte handelt. Kritiker merken an (Rindermann, 2006), dass mit der Messung kognitiver Fähigkeiten ohnehin das gleiche Konstrukt wie mit der Messung mathematischer Schulleistungen erfasst wird und insofern kaum Unterschiede deutlich werden können. Diese Ansicht wird von den Forschern abgelehnt, die von der Unabhängigkeit der Konstrukte ausgehen, während eine dritte Gruppe für gemeinsame Überschneidungsbereiche zwischen kognitiven Fähigkeiten und mathematischen Schulleistungen plädiert. Die relevante, grundlegende theoretische Debatte wird in Kapitel 2.2.1 aufgenommen, in dem bislang kontroverse Überzeugungen hinsichtlich der präzisen Trennbarkeit der Konstrukte mathematische Schulleistungen und kognitive Fähigkeiten diskutiert werden. Aufgrund der Annahme bereichsspezifischer kognitiver Fähigkeiten und stoffgebietsspezifischer mathematischer Fähigkeiten (Thurstone, 1938; Jäger et al., 1997; Perleth, 1997; Brunner et al., 2006) wird daher in der ersten Fragestellung thematisiert, inwiefern sich eine faktorielle Differenziertheit beider Konstrukte identifizieren lässt.

Die zweite Forschungsfrage in Kapitel 2.2.2 stellt die interne Vergleichbarkeit der Konstruktmessungen zwischen den Testverfahren selbst über mehrere Messzeitpunkte im Längsschnitt von vier Jahren in den Mittelpunkt. Bisher verwendete Tests sind zur querschnittlichen Erfassung gut geeignet (Krajewski et al., 2002, 2004; Roick et al., 2004, 2006). Ihre Konstruktkonsistenz für längsschnittliche Aussagen wurde bislang jedoch nur bedingt geprüft. Wird beispielsweise in der vierten Klassenstufe noch dasselbe Konstrukt erfasst, wenn Sachrechnungen mit Sachaufgaben der ersten Klassenstufe verglichen werden? Aufgrund der strukturellen Veränderung im Kindesalter und der quantitativen Steigerung kognitiver Leistungen (Hofstätter, 1954; Ziegler, 2004; Schneider & Hasselhorn, 2007; Rost, 2010) sind also Testadaptionen über den Messzeitraum nötig. Die sich daraus ergebenden Konsequenzen sind ebenfalls Gegenstand der Betrachtung.

Die dritte Forschungsfrage widmet sich der Stabilität kognitiver Fähigkeiten und mathematischer Schulleistungen. Allgemein wird von einer relativen Stabilität schulischer Leistungen ausgegangen. Diese Annahme bildet häufig schon in der Grundschule die Basis für frühe Schullaufbahnentscheidungen und Studienerfolg (Stumpf & Schneider, 2010). Doch stehen bei genauerer Recherche der relevanten Forschungsergebnisse nicht allzu viele empirische Belege mit tatsächlich belastbaren Messergebnissen für diesen Zeitraum zur Verfügung, sodass diese Annahme kritisch zu hinterfragen ist. Bisherige Befunde zur Stabilität kognitiver Fähigkeiten und mathematischer Schulleistungen werden daher in Kapitel 2.2.3 thematisiert. Es wird Bezug auf die Hauptdeterminanten mathematischer Schulleistungen genommen und auch entwicklungspsychologische Aspekte werden angesprochen.

Die vierte Fragestellung in Kapitel 2.2.4 wendet sich der gegenseitigen Beeinflussung kognitiver Fähigkeiten und mathematischer Schulleistungen zu. Dabei liegt das Augenmerk zunächst auf den Determinanten kognitiver Fähigkeiten und mathematischer Schulleistungen, denn die empirische Befundlage, deren Eckpunkte sich in diesem Zusammenhang im Zentrum des Interesses finden, ist hinsichtlich der möglichen Einflussfaktoren nicht eindeutig (Ceci, 1991; Watkins et al., 2007). Vor allem auch das Ausmaß, inwiefern kognitive Fähigkeiten Lernerfolge und Wissenserwerbsprozesse bestimmen oder aber vornehmlich wissensbasierte Faktoren spätere kognitive Fähigkeiten beeinflussen, ist eine in der Forschung noch nicht hinreichend beantwortete Frage. Neben den Wirkmechanismen kognitiver Fähigkeiten auf mathematische Schulleistungen gilt das Augenmerk in diesem Kapitel auch den umgekehrten Wirkmechanismen mathematischer Schulleistungen auf kognitive Fähigkeiten, während die interessierenden gegenseitigen Effekte ebenfalls erörtert werden. Allerdings bleibt die Frage empirisch zunächst unbeantwortet, da längsschnittliche

Studien bislang häufig sehr lange Zeiträume in das Untersuchungsdesign aufnehmen und daher die Ergebnisse keine zuverlässigen Aussagen für das Grundschulalter gestatten. Es gilt also, einen empirischen Nachweis zu Stärke und Richtung der gegenseitigen Beeinflussung kognitiver Fähigkeiten und mathematischer Schulleistungen im Grundschulalter zu erbringen, da Untersuchungen, die sich dieser Frage widmeten, bislang nur zwei Messzeitpunkte berücksichtigten (Watkins et al., 2007).

Der jeweils erarbeitete Sachstand zu den vier Fragen der faktoriellen Differenziertheit, der Invarianz, der Stabilität und der kreuzverzögerten Effekte erlaubt im jeweils direkten Anschluss nach einer kurzen Zusammenfassung die Ableitung der relevanten Forschungsfragen sowie die Formulierung der Hypothesen. So erfolgt im Kapitel 2.2.1.3 die Hypothesenformulierung zur ersten Forschungsfrage hinsichtlich der Beantwortung der Konstruktunabhängigkeit, während die Formulierung der Hypothesen zur Invarianz der Testverfahren zur Erfassung mathematischer Schulleistungen sowie der Invarianz der Testverfahren zur Erfassung kognitiver Fähigkeiten (zweite Forschungsfrage) in Kapitel 2.2.2.1 folgt. Die Hypothesenformulierung der dritten Forschungsfrage zur Stabilität kognitiver Fähigkeiten und mathematischer Schulleistungen über einen Vierjahreszeitraum im Grundschulalter erfolgt in Kapitel 2.2.3.3. Die Formulierung der Hypothesen zur vierten Forschungsfrage, die Einflüsse kreuzverzögerter Effekte kognitiver Fähigkeiten und mathematischer Schulleistungen der empirischen Überprüfung zugänglich machen sollen, erfolgt in Kapitel 2.2.4.2.

Im empirischen Teil der Arbeit, der zu Beginn den theoretischen Hintergrund unter Berücksichtigung der vier Forschungsfragen kurz rekapituliert, widmet sich die Arbeit dann deren empirischer Überprüfung in den Kapiteln 3 und 4. Um präzise Erkenntnisse über die komplexen Zusammenhänge zwischen kognitiven Fähigkeiten und mathematischen Schulleistungen zu gewinnen, werden beide Konstrukte über einen im Längsschnitt betrachteten Zeitraum von vier Jahren für das Grundschulalter betrachtet. Detaillierte Informationen zur Stichprobengrundlage – Daten einer Totalerhebung im Rahmen der Mecklenburger Längsschnittstudie aus den Jahren 2006 bis 2009 – werden in dem Kapitel 3.1 gegeben. Die relevanten Messinstrumente stehen im Kapitel 3.3 im Mittelpunkt. Diese standardisierten Testverfahren, die geeignet sind, kognitive Fähigkeiten sowie mathematische Schulleistungen zu erfassen, werden in den Kapiteln 3.3.1 und 3.3.2 skizziert. Anschließend wird ihre Eignung hinsichtlich der *längsschnittlichen Erfassung* der entsprechenden Fähigkeit erörtert. Informationen zum Studiendesign erfolgen im Kapitel 3.2, während in dem folgenden Ka-

pitel 3.4 die Analysemethoden vorgestellt werden. Die zur Beantwortung der Forschungsfragen eingesetzten Angaben zu den eingesetzten Analyseinstrumenten – die Softwareprogramme SPSS 20 und die Analysesoftware MPlus 7.1 und Analysestrategien erfolgen ebenso in diesem Kapitel. Diese empirischen Prüfverfahren werden im Detail dokumentiert, um zum einen das Analysedesign selbst und zum anderen das Verständnis in der Abfolge der Analysen transparent zu machen, da die Modellspezifikationen einem schrittweise aufeinander aufbauenden Vorgehen folgen.

Die sich anschließenden Kapitel widmen sich den empirischen Überprüfungen der formulierten Hypothesen.

Kapitel 4.1 beinhaltet Analysen und Ergebnisse hinsichtlich der Fragestellung 1. Hier wird überprüft, ob tatsächlich präzise trennbare Konstrukte erfasst werden oder ob beide Konstrukte vielmehr identische Fähigkeiten erfassen bzw. inwiefern Überlappungsbereiche existieren. In den Kapiteln 4.2 und 4.3 erfolgt die empirische Prüfung der zweiten Forschungsfrage zur Vergleichbarkeitsüberprüfung (Invarianz) der standardisierten Testverfahren für mathematische Schulleistungen (DEMAT) und kognitive Fähigkeitstests (KFT). Kapitel 4.4 beantwortet Fragestellung 3 zur Stabilität mathematischer Schulleistungen und der Stabilität kognitiver Fähigkeiten über den betrachteten Vierjahreszeitraum im Grundschulalter. Kapitel 4.5 dient der Beantwortung der vierten Fragestellung, in welcher der Frage der gegenseitigen Einflüsse kognitiver Fähigkeiten und mathematischer Schulleistungen nachgegangen wird. Dieses Kapitel umfasst Analysen und Ergebnisse kreuzverzögerter Effekte. In einem schrittweisen Modellaufbau erfolgt die Analyse der Einflüsse zunächst von kognitiven Fähigkeiten auf mathematische Schulleistungen aus vorherigen Messzeitpunkten auf Folgejahre. Anschließend werden die Einflüsse in umgekehrter Richtung – von früheren mathematischen Schulleistungen auf spätere kognitive Fähigkeiten – untersucht, während in der Folge die Modellspezifikation insofern erweitert wird, als dass nunmehr die gegenseitigen zeitverzögerten Effekte von früheren Leistungen eines Konstrukts auf spätere Leistungen des jeweils anderen Konstrukts gleichzeitig erfasst werden.

Im Kapitel 5 erfolgt die Gesamtdiskussion und die Untersuchungsergebnisse werden in Beziehung zur theoretischen Basis gesetzt. Implikationen und Perspektiven für die Forschungs- und pädagogische Praxis bilden den Schlussteil.

2 Theoretischer Teil

2.1 Kognitive Fähigkeiten und mathematische Schulleistungen im Grundschulalter

Im Fokus der Studie steht das Anliegen, präzise Aussagen über den Zusammenhang von kognitiven Fähigkeiten und mathematischen Schulleistungen machen zu können. Dazu müssen zunächst beide Konstrukte definiert werden.

In einem ersten Schritt soll das Konstrukt kognitive Fähigkeiten definiert werden. Da kognitive Fähigkeiten als maßgeblich für die erfolgreiche Bewältigung spezifischer und allgemeiner Anforderungen im Umgang mit zahlreichen unvorhersehbaren Problemen des Lebens gelten (Carroll, 1963, 1989, 1993), kommt der Verwendung des Begriffs eine enorme Bedeutung zu. In der Auseinandersetzung mit dem Terminus kognitive Fähigkeiten wird allerdings offensichtlich, dass es keine einheitliche Definition gibt. Zudem werden die Begriffe kognitive Fähigkeiten, allgemeine Denkfähigkeit oder auch Intelligenz häufig synonym verwendet. Als Indiz für die Schwierigkeiten, die mit der Definition von Intelligenz und ihrer Abgrenzung zum Begriff kognitive Fähigkeiten verbunden sind, mag der Hinweis gelten, dass es etwa 3500 verschiedene Intelligenztests mit häufig unterschiedlichen Definitionen gibt. Es ist daher geboten, zunächst einen Blick auf das Konzept Intelligenz zu werfen, um sich im Anschluss einer Konkretisierung des Begriffs kognitive Fähigkeiten, wie er im Rahmen dieser Arbeit zu definieren ist, anzunähern. Relevante Theorien und Forschungsbefunde zu den Konzepten und ihrer Entwicklung stehen anschließend im Zentrum des Interesses.

Der Definition des Begriffs mathematische Schulleistungen mit der Skizzierung entsprechender Modelle kommt ein ebenso hoher Stellenwert zu. Sie besitzen für den schulischen Kontext einen mindestens ebenso großen Einfluss wie kognitive Fähigkeiten, denn sie werden als eine wesentliche Determinante des Schulerfolgs betrachtet. Außerdem sind mathematische Fähigkeiten Basis naturwissenschaftlicher Schulfächer und werden vielfach in Auswahlverfahren für Berufs- und Studienzulassungen abverlangt.

In diesem Zusammenhang ist es zur Eingrenzung des Terminus notwendig, zunächst kurz auf die Begriffe Leistung im Allgemeinen, Schulleistungen im Besonderen und schließlich mathematische Schulleistungen im ganz Speziellen einzugehen. Die standardisierte Erfassung beider Konstrukte steht insbesondere unter dem Aspekt der stoffgebiets- bzw. bereichsspezifischen Erfassung der Faktoren im Zentrum. Hinsichtlich der präzisen Trennbarkeit der Konstrukte mathematische Schulleistungen und kognitive Fähigkeiten

werden die Theoriekontroversen aufgenommen und diskutiert, welche die Basis zur Formulierung der Fragestellung 1 bilden. Vor dem Hintergrund der längsschnittlichen Studienanlage dieser Arbeit wird besonderer Wert auf die Konstruktkonsistenz der Testverfahren gelegt und der theoretische Zugang hinsichtlich der Fragestellung 2 in Kapitel zur Invarianzprüfung dargelegt. Die Entwicklung kognitiver Fähigkeiten und mathematischer Schulleistungen im Grundschulalter, wie die Fähigkeit zur niveaugesteigerten Informationsverarbeitung und steigenden Problemlösekompetenzen, bilden die theoretische Grundlage zur Formulierung der Stabilitätsfrage (Fragestellung 3) kognitiver Fähigkeiten und mathematischer Schulleistungen. Der theoretische Forschungsstand zu den wechselseitigen Einflüssen mathematischer Schulleistungen und kognitiver Fähigkeiten wird anschließend in Fragestellung 4 diskutiert.

2.1.1 Kognitive Fähigkeiten

Die Konzeptualisierung der Intelligenz zieht seit über einem Jahrhundert große Aufmerksamkeit auf sich. Charles Spearman postulierte schon 1904 die Existenz eines sogenannten Generalfaktors g , der nach seinen Vorstellungen die Allgemeine Fähigkeit zum Denken definierte. William Stern seinerseits erklärte im Jahr 1912 Intelligenz als die allgemeine Fähigkeit eines Individuums, sein Denken bewusst auf neue Forderungen einzustellen. Er sah die Intelligenz als allgemeine geistige Anpassungsfähigkeit an neue Aufgaben und Bedingungen des Lebens an (Stern, 1912; in Heller, 1976).

Doch bereits früh in der Entwicklung der Intelligenzforschung wurde deutlich, dass über diese eingeschränkte Definition von Intelligenz als einer allgemeinen Fähigkeit zum Denken keine Einigkeit zu erzielen war. Im Jahr 1921 scheiterten die führenden Forscher ihrer Zeit, als sie sich bei einem von den Herausgebern des *Journal of Pedagogic Psychology* einberufenen Symposium an einer einheitlichen Definition von Intelligenz und ihrer Messung versuchten (Schmiedek & Li, 2015). 14 der 17 Teilnehmer vertraten unterschiedliche Ansichten. Ein Viertel der teilnehmenden Forscher sprach sich für elementare sensorische und kognitive Prozesse (Wahrnehmung und Aufmerksamkeit) als primäre Attribute der Intelligenz aus. Ein weiteres Viertel der Experten war der Ansicht, dass die physiologischen Mechanismen des Gehirns die bestimmenden Faktoren der Intelligenz sind, andere definierten adaptive Verhaltensreaktionen als Schlüsselmerkmale der Intelligenz. Fünfundsechzig Jahre später herrschte noch der gleiche Mangel an Konsens. Sternberg und Detterman trugen im Jahr 1986 die verschiedenen existierenden Definitionen zusammen und präsentierten

auf einem Nachfolge-Symposium mit 27 Experten eine lange Liste mit Definitionen von Intelligenz und überlappenden Komponenten (Sternberg & Dettermann, 1986).

Wie schwierig es auch im neuen Jahrtausend noch immer ist, Intelligenz zu definieren, zeigten beispielsweise Wasserman und Tulsky, als sie im Jahr 2005 eine Liste verschiedener Definitionen veröffentlichten. Eine ergänzende Liste mit 19 weiteren verschiedenen Definitionen, die von dominierenden Experten in der Psychologie anerkannt worden war, wurde im Jahre 2008 veröffentlicht. Dass das Kapitel der Definitionsfindung längst nicht abgeschlossen ist, betont Wasserman im Jahr 2012 erneut (Wasserman & Tulky, 2005, 2012). Ein Versuch, diese zahlreichen unterschiedlichen Definitionen nach übereinstimmenden Elementen zu untersuchen, wurde im Jahr 1987 von Snyderman und Rothman (1987) unternommen. Sie konnten drei Definitionselemente ermitteln, über die zumeist Konsens besteht. Diese drei Elemente sind a) abstraktes Denken, b) die Fähigkeit, Wissen zu erwerben, und c) die Problemlösungsfähigkeit (Kaya et al., 2015).

Auch unstrittig ist, dass Intelligenz, verstanden als die allgemeine Fähigkeit zum schlussfolgernden Denken und zum Abstrahieren sowie als das Verständnis von theoretischen Zusammenhängen, jedoch nicht alle Dimensionen erfassen kann, die das Konstrukt auszeichnen. Als weitere wichtige zugrundeliegende Komponenten der Intelligenz gelten zum Beispiel neuronale Elemente, wie die Informationsgeschwindigkeit oder auch die Prozessgeschwindigkeit (Mackintosh, 2011). Aber auch diese Ergänzungen liefern noch keine hinreichende Erklärung. In aktuellen Diskussionen wird beispielsweise immer wieder auf die Berücksichtigung von Wissen und Vorwissenselementen (Ackerman, 2011; Perleth, 1997; Rindermann, 2006; Ceci, 1991; Heller, 1991; Krajewski 2008), oder auf die Bedeutung des erfolgreichen Abrufs relevanter Gedächtnisinhalte (Gruber, 2008; Ackerman et al., 2005) bzw. auf Lern- und Übungskonzepte (Ericsson, Krampe & Tesch-Römer, 1993) aufmerksam gemacht.

Wesentliche weitere Diskussionsstränge entstehen darüber, ob Intelligenz ein- oder multidimensional zu verstehen ist (Davis et al., 2011). Lebhaft diskutiert wird auch, ob Intelligenz als statisches oder dynamisches Konstrukt unterschieden werden soll oder nicht. Ein weiterer Hauptpunkt der Auseinandersetzungen ist, ob Intelligenz als globale intellektuelle Fähigkeit oder als Anordnung separater Fähigkeiten verstanden wird. Eine Reihe von Forschern betont, dass die Bereichsspezifität von Intelligenz betrachtet werden müsse. Beispielsweise unterschied Thurstone (1931, 1938) sieben, später neun Primärfaktoren der In-

telligenz. Ihm folgten Cattell (1963) mit der bekannten Zwei-Faktoren-Theorie kristallisierter und fluider Intelligenzvorstellungen und Gardner (1983) mit der Konzeption multipler Intelligenzen. Auf diese Konzepte wird im Laufe der Arbeit zurückzukommen sein.

Zusammenfassend lässt sich festhalten, dass in Expertengruppen trotz unzähliger Versuche, eine einheitliche Definition von Intelligenz zu formulieren, bis heute keine Einigung zu erzielen war. Es werden sowohl grundlegende kognitive Prozesse, metakognitive Prozesse, Anpassung an die Umwelt, Argumentationsfähigkeit, Problemlösung und Entscheidungsfindung diskutiert (Kaya, 2015). Zahlreiche Definitionen existieren nebeneinander, die je nach dem Modell der Intelligenz, auf das ein Forscher sich bezieht, differieren. Die zugrundeliegenden Konzeptionen unterscheiden sich im Wesentlichen dahingehend, ob sie einen generellen Faktor der Intelligenz annehmen oder nicht (Mackintosh, 2011). Wird nicht grundsätzlich von einer allgemeinen Fähigkeit ausgegangen, so werden spezifische Fähigkeiten für Intelligenzbereiche angenommen. Diese werden zum Teil nebeneinander separat stehend betrachtet, oder aber es wird ein hierarchischer Aufbau angenommen.

Dem Vorschlag Gottfredssons (1997), dieser Vielzahl an Definitionen ein Ende zu setzen und eine Mainstream-Definition von Intelligenz zu finden, folgten schließlich 52 Intelligenzforscher (Kaya, 2015). Nach dieser Definition ist Folgendes unter Intelligenz zu verstehen:

„Intelligenz ist die allgemeine mentale Fähigkeit, die unter anderem die Fähigkeit beinhaltet, zu schlussfolgern, zu planen, Probleme zu lösen, abstrakt zu denken, komplexe Ideen zu verstehen, schnell und aus Erfahrung zu lernen. Sie sei nicht nur eine akademische Fähigkeit, Tests erfolgreich zu lösen, sondern vielmehr eine breit zu verstehende Fähigkeit, zugrundeliegende Hintergründe zu begreifen, die Umgebung zu verstehen, ein Sinnverständnis theoretischer Zusammenhänge zu entwickeln und weiter, die Fähigkeit zur Entscheidungsfindung.“ (Gottfredsson, 1997)

Diese Definition kann somit als der kleinste gemeinsame Nenner, auf den sich Forschergruppen geeinigt haben, verstanden werden. Darüber hinaus bestehen kaum einheitliche Vorstellungen. Das oben beschriebene Verständnis von Intelligenz impliziert einerseits eine relativ scharfe Abgrenzung, postuliert aber andererseits vielfältige Bereiche intellektueller Fähigkeiten, also eigentlich Intelligenzen im Plural. Da diese Begriffsverwendung jedoch Irritationen hervorrufen würde, setzte sich der Begriff kognitive Fähigkeiten durch, den Carroll (1993) schlicht als Potential, eine definierte Klasse von Aufgaben erfolgreich zu

bearbeiten, definierte. Die Verwendung des Begriffs Fähigkeiten im Plural trägt der Tatsache Rechnung, dass Fähigkeiten vielfältig und damit zugleich bereichsorientiert sein können. Auf diese Weise wird auch der Dynamik der Veränderlichkeit dieser Fähigkeiten Rechnung getragen. Insofern werden kognitive Fähigkeiten als dynamisch und multidimensional sowie als Potential und Ressource verstanden (Perleth, 1997). Da der Begriff Fähigkeit als Potential verstanden wird und somit impliziert, dass diese Fähigkeit erweiterbar oder auch vertiefbar ist, kann mit dem Begriff kognitive Fähigkeiten ein breiteres Verständnis als mit dem Begriff Intelligenz ausgedrückt werden. Aus diesen Gründen wird in der vorliegenden Arbeit der Begriff kognitive Fähigkeiten verwendet und zugleich einer synonymen Verwendung mit dem Intelligenzbegriff zugestimmt.

Intelligenz im Sinne kognitiver Fähigkeiten soll also im Rahmen dieser Arbeit als die „...komplexe Problemlösefähigkeit bei sich ändernden Anforderungen und (konkreter), als Maß, wie wir Umweltanforderungen bewältigen“, definiert werden (Wittmann, Eisenkolb & Perleth, 1997).

Die ausgewählte Definition von Wittmann, Eisenkolb und Perleth (1997) folgt der Tradition Thurstones (1931, 1938), der von speziellen kognitiven Fähigkeiten ausging. Die Annahme von spezifischen kognitiven Fähigkeiten soll im Rahmen der vorliegenden Arbeit theoretisch und empirisch begründet werden.

2.1.1.1 Forschungstraditionen in der Strukturanalyse von kognitiver Fähigkeiten

Im Folgenden soll nun der Blick auf die für diese Arbeit wichtigsten Forschungstraditionen sowie Intelligenztheorien und ihre Entwicklung gelenkt werden, um sich dem Begriff kognitive Fähigkeiten und seiner Bedeutung für die vier der Arbeit zugrundeliegenden Fragen weiter anzunähern.

McGrew und Flanagan (1997) identifizieren drei verschiedene Forschungstraditionen in der Strukturanalyse von Intelligenz. Hierbei handelt es sich zum einen um den Informationsverarbeitungsansatz, der computerbasierte Analysemethoden verwendet, um kognitive Informationsprozesse in der Auseinandersetzung mit problemorientierten Anforderungen zu offenbaren. Zum anderen weisen sie auf den Ansatz der kognitiven Modifizierbarkeit hin, der auf dem anpassungsfähigen und veränderlichen Charakter der Intelligenz basiert und sich auf die dynamische Veränderung konzentriert. Beide Ansätze bleiben in ihrer Bedeutung hinter dem dritten Ansatz zurück. Hierbei handelt es sich um den psychometrischen

bzw. strukturellen Ansatz, der auf Datenreduktionstechniken wie der Faktorenanalyse beruht. Er ist der älteste und zugleich dominierende Ansatz innerhalb der Forschungstraditionen in der Strukturanalyse von Intelligenz (Neisser et al., 1996) und grundlegend für diese Arbeit. Daher sollen jetzt ausgewählte psychometrische Intelligenztheorien kurz dargestellt werden.

Die Theorie von Spearman

Nachdem Spearman zunächst mit der Etablierung von g die Ein-Faktor-Theorie begründet hatte (1904), modifizierte er im Laufe seiner bahnbrechenden Forschungen diese Theorie der allgemeinen Intelligenz und benannte sie nunmehr in Zwei-Faktor-Theorie um. Die zugrundeliegende Annahme des neuen Modells ist, dass individuelle Unterschiede in zwei Komponenten zerlegt werden können. Eine Komponente g ist an dem Abschneiden aller kognitiven Aufgaben beteiligt, während spezifische Komponenten für jede Aufgabe unabhängig sind (Spearman, 1927).

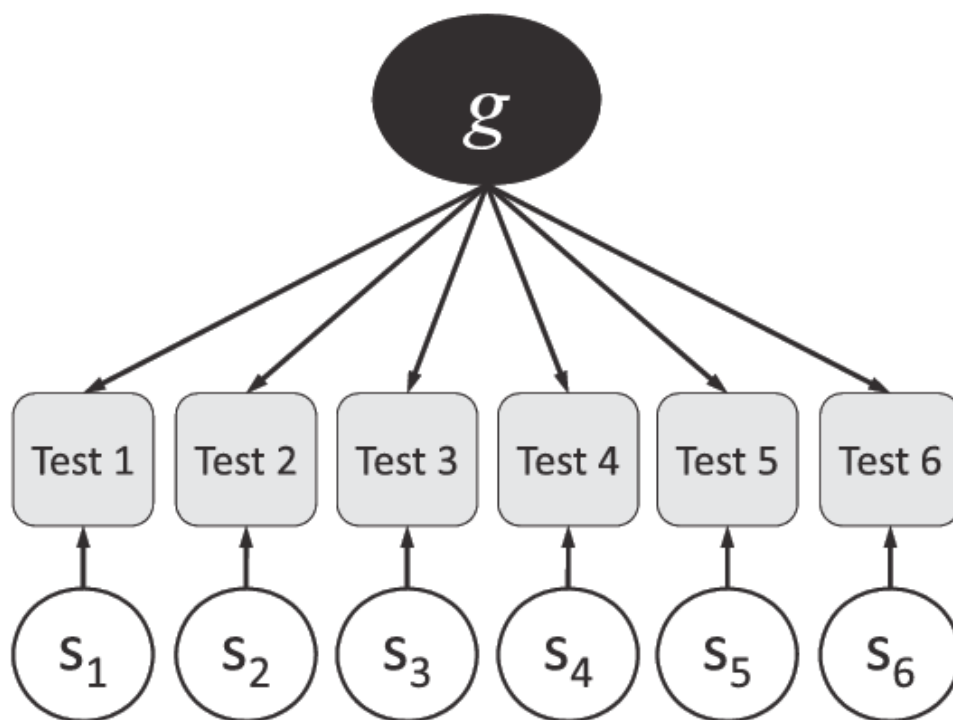


Abbildung 1 Spearmans g (aus Schneider & McGrew, 2012).

In diesem Zwei-Faktoren-Modell akzeptierte er nun die im Leistungszusammenhang als relevant erachteten spezifischen Faktoren (Spearman, 1927, In: Schneider & McGrew, 2012).

Die Spearman'sche Zwei-Faktoren-Theorie wurde weiterentwickelt und es wurden Gruppenfaktoren extrahiert (Willis et al., 2011). Spearmans „g-Theorie“ dominierte für mehr als ein dreiviertel Jahrhundert die Forschungen. Die Frage, ob die Intelligenz hauptsächlich ein allgemeiner Faktor im Sinne von Spearmans g ist, stand dabei im Mittelpunkt des Interesses. Forscher wie Thorndike et al. (1926) gingen wiederum früh der Frage der Bereichsspezifität der Intelligenz nach. Sie argumentierten, dass der allgemeine Faktor mit verschiedenen spezifischen Fähigkeiten interagiert und sprachen sich für eine Überlappung eines generellen Faktors g mit verschiedenen Subfähigkeiten aus.

Die Theorie von Thurstone

Auf Spearmans Theorie folgte Thurstones Modell der Intelligenz (1938). Thurstone ging dabei nicht von einem übergeordneten allgemeinen Intelligenzfaktor aus, sondern separierte sieben, später neun sogenannte Primärfaktoren, die er als voneinander unabhängig erkannte. Die zunächst von Thurstone extrahierten sieben Primärfaktoren sind:

- "verbal comprehension" (Erfassung sprachlicher Bedeutungen und Beziehungen sowie Umgang mit sprachlichen Begriffen)
- "word fluency" (Wortflüssigkeit und assoziative Wortproduktion)
- „memory“ (Gedächtnis, Kurzzeitgedächtnis, Merkfähigkeit)
- "reasoning" (Denkfähigkeit zum schlussfolgernden Denken, induktive Denkfähigkeit und Regelanwendung)
- "numbers" (Rechenoperationen)
- "space" (räumliches Vorstellungsvermögen)
- "perceptual speed" (Wahrnehmungsgeschwindigkeit und wahrnehmungsgebundenes logisches Denkvermögen)

Thurstones Theorie der Intelligenz (1938) stellte im Gegensatz zu Spearmans Theorie nicht dichotome Fähigkeiten dar, sondern erklärte Intelligenz durch primäre geistige Fähigkeiten. Im Gegensatz zu der Annahme eines generellen dominierenden g-Faktors werden Thurstones primäre geistige Fähigkeiten als alle gleichermaßen wichtig erachtet. Der Spearman'sche g-Faktor spielte eine untergeordnete Rolle bei Thurstones Erklärung der Intelligenz. Er stand also zunächst im Gegensatz zur Spearman'schen Theorie, doch wurde seine Annahme dieser primären geistigen Fähigkeiten vollständig orthogonal zueinanderstehender Faktoren nicht empirisch unterstützt. Deshalb akzeptierte er später vorsichtig die Existenz eines allgemeinen g-Faktors. Wasserman und Tulsky (2005) sehen Spearman und

Thurstone deshalb als zwei Extreme eines Kontinuums: Spearman schlägt einen allgemeinen Intelligenzfaktor vor, während Thurstone sieben separate primäre Fähigkeiten favorisiert. Die endgültige Formulierung dieser Modelle brachte sie jedoch konzeptuell näher. Spearmans Akzeptanz spezifischer Faktoren zusätzlich zu g und Thurstones spätere Akzeptanz der Anwesenheit eines allgemeinen g-Faktors zusätzlich zu seinen sieben primären Fähigkeiten näherte nicht nur diese Theorien einander an, sondern bot auch eine Grundlage für spätere Forschungen.

Vernon (1950) plädierte in der Folge für eine hierarchische Gruppierung der Intelligenz. Er begründete damit die Faktoranalyse. In seiner Konzeption steht auf der obersten Ebene eine generelle Fähigkeit. Spezifische Faktoren wie verbale Leistungen und räumliche mechanische Fähigkeiten werden dagegen als weniger dominierend angesehen. Faktoren wie verbale Wortflüssigkeit, numerische Fähigkeiten oder psychomotorische Fähigkeiten werden in der Hierarchie niedriger angeordnet. Seine Annahme bereichsspezifischer Fähigkeiten unter Akzeptanz eines übergeordneten g-Faktors erreichte weitreichende Bedeutung. Zu betonen ist beispielsweise die explizite Extraktion numerischer Fähigkeiten, die nach Thurstone ausdrücklich von anderen Fähigkeiten abgegrenzt werden.

Die Cattell-Theorie unter gleichzeitiger Berücksichtigung seiner Investitionstheorie

Raymond Bernard Cattell (1963) nahm Thurstones Primärfaktorenmodell in sein Konzept der Intelligenz, das er in den Jahren 1957, 1963, 1965, 1971 und 1974 entwickelte und weiterentwickelte, auf und maß den erwähnten erworbenen Fähigkeiten eine enorme Bedeutung bei. Er bezeichnete sie als kristallisierte Intelligenzanteile (crystallized g, auch gc), welche gemeinsam mit den allgemeinen Denkfähigkeiten im Sinne Spearmans, die er als fluide Intelligenzanteile (fluid g, auch gf) bezeichnete, dazu führen, dass komplexe Anforderungen erfolgreich bewältigt werden können. In diesem Modell fluider und kristallisierter Intelligenz ist ein allgemeiner Intelligenzfaktor g auf der ersten Ebene organisiert. Weitere Faktoren werden auf der zweiten Ebene hinzugefügt, während spezifische Faktoren auf der dritten Ebene hinzukommen.

Die Interaktion fluider und kristallisierter Intelligenz wird von Cattell (1971) als ontogenetische Investitionstheorie bezeichnet. Er postuliert, dass die kultur-, erfahrungs- und wissensbasierten kristallisierten Fähigkeiten zum Teil als Folge von biologischen reifungs- und prozessbasierten fluiden Fähigkeiten entstehen, wenn diese in geeignete Lernerfahrungen und Kontexte investiert werden.

Fluide Intelligenz (g_f) besitzt daher vor allem in frühen Phasen des Lernens und in neuartigen Situationen besondere Bedeutung, wenn sich Lernende mit neuen Informationen oder Erfahrungen konfrontiert sehen, die zunächst als unorganisiert empfunden werden. Sie unterstützt die zügige und effektive Aneignung von Fähigkeiten und Kenntnissen (Voelkle, Wittmann & Ackerman, 2006; Watkins et al., 2007). Das Fokussieren auf neuartige Situationen deutet an, dass intelligentes Verhalten hier mit der Fähigkeit, sich schnell und erfolgreich mit unbekannten Problemstellungen auseinandersetzen zu können, assoziiert wird. Der Faktor Schnelligkeit deutet an, dass für verschiedene Problemstellungen nicht unbegrenzt Zeit zur Verfügung steht. Fluide Intelligenz erlaubt einem Individuum allgemeines Denken, sowie schnelle Reaktionen und angemessene Lösungen in der Konfrontation mit neuartigen Problemen. Ein gut organisiertes Kurzzeitgedächtnis unterstützt dies. Cattells (1987) Hypothesen sprechen für die enge Beziehung zwischen kognitiven Leistungen und schulischen bzw. akademischen Leistungsfaktoren. Nach der Investitionstheorie wird fluide Intelligenz als eine der Hauptursachen für die Leistung angesehen, da sie die Kapazität zum schnellen Lernen bedeutet (Lemos et al., 2011).

Die kristallisierte Intelligenz wird dagegen von Cattell als tiefes und breites, kulturell geschätztes Wissen definiert (Cattell, 1971; Schneider & McGrew, 2012). Kristallisierte Intelligenz umfasst hauptsächlich erworbenes Wissen in der Problemlösung und erfordert die Verfügbarkeit von Informationen und Sprache.

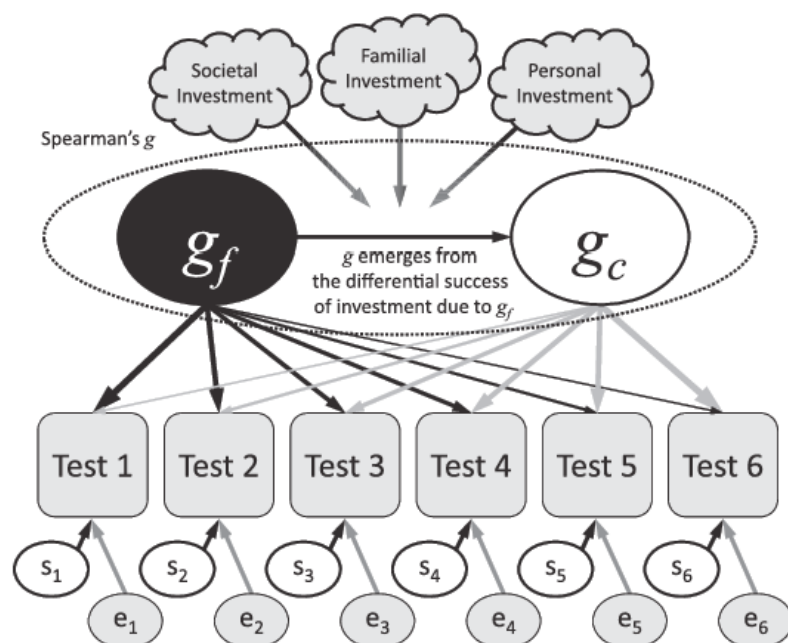


Abbildung 2 Cattell's investment theory (aus Schneider & McGrew, 2012).

Zusammenfassend lässt sich sagen, dass Cattell gleichsam Sozialisationseinflüsse integriert, indem er das erworbene Wissen einer Person, welches wiederum abhängig von Umwelteinflüssen und Lerngelegenheit ist, in den Rahmen der kristallisierten Intelligenz einfügt. Damit subsumiert er die kristallisierte Intelligenz als erworbenes Wissen und stellt die Bedeutung der Lernumwelt heraus. Die Cattells Theorie zugrundeliegenden Ideen etablierten theoretische Vorstellungen von höher organisierten, also hierarchisch angeordneten, kognitiven Fähigkeiten, die abgegrenzt von mehr oder weniger unabhängigen Fähigkeiten angenommen werden (Cattell & Horn, 1978).

Cattells Theorie leitete eine neue Entwicklungsstufe moderner multipler Intelligenzstrukturtheorien ein, bei denen Cattells Konzeption wegweisend blieb und auch aktuell weitgehende Akzeptanz findet. So definieren Lemos et al. (2011) Intelligenz in Anlehnung an Cattells fluide Intelligenz (1971) als induktives und deduktives Denkvermögen. Es beinhaltet die Fähigkeit zu denken, Probleme zu lösen, Schlüsse zu ziehen, Beziehungen zu identifizieren sowie Informationen in relevanter Weise zu verknüpfen (Lemos, Almeida & Colom, 2011). Auch Kvist und Gustafsson (2008), Voelkle, Wittmann und Ackerman (2006) sowie Watkins, Lei und Canivez (2007) sehen die fluide Intelligenz (*gf*) im Sinne Cattells als Grundlage kristallisierter Intelligenz (*gc*), da fluide Intelligenz den Erwerb von Fähigkeiten und Wissen im Sinne kristallisierter Intelligenz mindestens unterstützt, wenn nicht erst ermöglicht. Primi (2010) schließt sich diesem Verständnis an und definiert fluide Intelligenz in der psychometrischen Tradition als die Verwendung von bewussten geistigen Operationen in der Auseinandersetzung mit neuen Problemen. Er schließt explizit Aufgaben, die als Funktion einer einfachen Speicherung oder Routine gelöst werden können, von der Definition fluider Intelligenz aus. Integrierte Operationen sind dagegen das Schlussfolgern, das Konzeptionalisieren, die Fähigkeit zur Klassifizierung, das Erzeugen und Prüfen von Hypothesen, das Identifizieren von Beziehungen, die Problemlösung und die Transformation von Informationen (McGrew, 2009).

Die Theorie von Cattell und Horn

Durch Weiterentwicklungen von Cattells Modell (1963) durch Cattell und Horn (1978) und nachfolgend von Horn und Noll (1997) wurde die Gültigkeit des Ausgangsmodells schließlich erheblich erweitert. Es entstand die einflussreiche Cattell-und-Horn-Theorie fluider und kristallisierter Intelligenz (*Gf-Gc*-Theorie). Sie ist die wahrscheinlich bekannteste und am weitesten verbreitete hierarchische Theorie der Intelligenz. Basierend auf Faktoranalyseergebnissen unterscheidet diese Theorie einerseits die allgemeinen fluiden Fähigkeiten (*Gf*),

in denen Fähigkeiten wie Problemlösung, Argumentation, Lernen und selektive Aufmerksamkeit abgebildet werden und andererseits die kristallisierten Fähigkeiten (*Gc*), die die erworbenen Kenntnisse abbilden. Flanagan und McGrew (1997) bestätigten, dass das *Gf-Gc*-Intelligenzmodell eines der umfassendsten empirisch validierten Theorien der Intelligenz ist. Fluide Intelligenz definieren Flanagan und McGrew daher in Anlehnung an die *Gf-Gc*-Theorie als die Verwendung von zielgerichteten und kontrollierten mentalen Operationen zur Lösung neuer Probleme, die nicht mittels Routinen oder als Ergebnis des Auswendiglernens gelöst werden können. Für die Repräsentation fluider mentaler Operationen stehen das Problemlösen, die Identifizierung von Beziehungen, das Verständnis von Konzepten, die Klassifizierung, das Schlussfolgern, die Transformation von Informationen sowie induktives und deduktives Denken (McGrew, 2009). Messick (1984) verglich ebenfalls zeitgenössische multidimensionale Modelle der Intelligenz und kam zu dem Schluss, dass die *Gf-Gc*-Theorie der Intelligenz im Vergleich mit anderen Theorien die umfassendsten Gültigkeitsnachweise liefert.

Auch aktuelle Forschungsbemühungen beziehen sich in ihren Untersuchungen auf die weithin akzeptierte Theorie, so beispielsweise Willis, Dumont und Kaufman (2011). Sie definieren fluide Intelligenz in Anlehnung an Cattell und Horn als induktive, deduktive und quantitative Begründungsfähigkeit mit Materialien und Verfahren, die neu für die Person sind. Fluide Fähigkeiten ermöglichen nach ihrer Definition individuelles Denken und schnelles Handeln, das Lösen neuer Probleme und das Abrufen kurzfristiger Erinnerungen. Die kristallisierte Intelligenz wiederum wird von Willis, Dumont und Kaufman (2011) in Anlehnung an Cattell und Horn als die Anwendung von erworbenen Kenntnissen und gelernten Fähigkeiten zur Beantwortung von Fragen zur Lösung von Problemen bezeichnet, die zumindest weitgehend vertraute Materialien und Prozesse darstellen. Sie spiegelt sich in Wissenstests unter Anwendung von Sprache bzw. Vokabular und einer Vielzahl von erworbenen Fähigkeiten wider. Im Zusammenhang mit der kristallisierten Intelligenz ist also das Lernen aus vergangenen Erfahrungen wesentlich. Beispiele für kristallisierte Intelligenz sind Leseverständnis oder Leistungen aus mathematischen Schulleistungstests.

Wenngleich Cattell und Horns Theorie fluider und kristallisierter Denkfähigkeiten weitgehende Akzeptanz genießt, blieb sie dennoch nicht unkritisiert. Alternative Konzepte zu der *Gf-Gc*-Theorie von Cattell und Horn gehen davon aus, dass Intelligenz aus den Bereichen verbale Fähigkeiten sowie perzeptuelle und mentale Vorstellungsbildung besteht, die einem übergeordneten *g*-Verständnis folgen. Auf einer dritten untergeordneten Ebene werden von Johnson und Bouchard unter anderem Zählfähigkeiten angeordnet (2005). Auch

müsste die ursprüngliche *Gf-Gc*-Theorie von Cattell und Horn nach Meinung vieler Experten mehr zwischen den Effekten kognitiver Fähigkeiten und denen der Schulbildung differenzieren, damit etwa die tatsächlichen wirtschaftlichen Ergebnisse beider Konstrukte als wesentlichste Wirkung betrachtet werden.

Stephen J. Ceci (1991) kritisiert Cattells Sichtweise kristallisierter Intelligenz eher als eine Reaktion auf Umwelteinflüsse, einschließlich der Schulbildung selbst. Auch die fluide Intelligenz vermuten Ceci und Williams stärker durch die neurologische Entwicklung beeinflusst (1997). Das muss ihrer Meinung nach durch umfangreiche statistische Beweise gestützt werden. Zudem thematisieren Ceci und Williams eine weitere offene empirische Frage. Obwohl Horn (1994) annimmt, dass die fluide Denkfähigkeit im Gegensatz zum kristallisierten Denkvermögen mit dem chronologischen Alter abnimmt, sprechen einige Untersuchungsergebnisse dagegen. Flynn berichtete beispielsweise von dem größten Zugewinn fluider Denkfähigkeiten bei Erwachsenen und niedrigeren Zuwächsen für die kristallisierten Anteile (1987). Auch Cahan und Cohen stützten diese These (1989). Die daraufhin von Ceci und Williams 1997 vorgeschlagene Ausdifferenzierung sah mehr als vierzig sogenannte Primärfaktoren vor und unterschied die kognitiven Fähigkeiten insgesamt in etwa zehn sogenannte verwandte Faktoren, die wiederum in die Unterscheidung der *Gf-Gc*-Theorie zweiter Ordnung einfließen. Dennoch wird auch von Ceci und Williams anerkannt, dass die Unterscheidung in die Faktoren *Gf* und *Gc* einen vernünftigen Kompromiss darstellt, obwohl intellektuelle Fähigkeiten durch fluide Fähigkeiten gestützt werden und ihrerseits die Schulbildung fördern. Insofern werfen Ceci und Williams der *Gf-Gc*-Theorie eine gewisse Unspezifität vor, da sie offen lasse, wie die tatsächlichen Pfade verlaufen, die die unterschiedlichen Wirkungen auf schulische und ökonomische Ergebnisse erzielen.

Kvist und Gustafsson (2008) sowie Watkins, Lei und Canivez (2007) dagegen vertreten in Übereinstimmung mit der Cattell-Investitionstheorie (Cattell, 1971) die Position, dass fluide Intelligenz den Erwerb von Fähigkeiten und Wissen, die das Wesen der kristallisierten Intelligenz sind, unterstützt. Fluide Intelligenz stellt daher eine Grundlage für kristallisierte Intelligenz dar und nicht eine, in einem *g*-Faktor aufgehende *allgemeine Intelligenz*. Sie gehen davon aus, dass *gf* und *gc* zwei verschiedene Konstrukte sind.

Diese zentralen Fragen, inwieweit kristallisierte Intelligenz identisch mit erworbenem Wissen ist, und inwiefern fluide Intelligenz gleichsam ein allgemeiner *g*-Faktor ist, werden im Rahmen der ersten Fragestellung dieser Arbeit thematisiert. Obwohl die Nützlichkeit der *Gf-Gc*-Theorie vielfach belegt ist, fällt doch auf, dass nicht alle Fragen zu den

Zusammenhängen beantwortet sind. Die genannten empirisch und konzeptionell noch offenen Fragen können anhand der Prüfung der gegenseitigen Einflüsse im Rahmen der vierten Fragestellung beantwortet werden.

Die Theorie von Carroll

Aufgrund der oben aufgeführten Kritik wurde die *Gf-Gc*-Theorie seither weiterentwickelt, zum Beispiel durch Überlegungen von Noll (1997). Es wurden weitere Faktoren hinzugefügt, etwa quantitatives Schlussfolgern (*Gq*), das Kurzzeitgedächtnis (*Gsm*), visuelle Verarbeitung (*Gv*), auditive Verarbeitung (*Ga*), der Langzeitabruf (*Glr*), die Verarbeitungsgeschwindigkeit (*Gs*), das Lesen und Schreiben (*Grw*) sowie die Reaktionsgeschwindigkeit (*CDS*). Horn und Noll (1997) behaupten, dass alle menschlichen Fähigkeiten durch die in dieser erweiterten Theorie enthaltenen Faktoren repräsentiert werden können.

John B. Carroll trug ebenfalls zu einer viel beachteten Erweiterung bei. Er untersuchte in einer groß angelegten Reanalyse alle zur Verfügung stehenden kognitiven Tests hinsichtlich ihrer Korrelation und fand in über 1500 Studien mittels Faktoranalysen die Domänen verbale Fähigkeiten, schlussfolgerndes Denken, Erinnerung, visuelles und auditives Lernen, Ideenproduktion, kognitive Geschwindigkeit, Wissenserwerb und Leistung, psychomotorische Fähigkeiten und Persönlichkeitsfaktoren (1993, 1997). Er präsentierte daraufhin seine Theorie der kognitiven Fähigkeiten, die Drei-Stratum-Theorie. Hier sind auf der ersten hierarchischen Ebene mehrere low-level-spezifische Fähigkeiten organisiert, die untereinander zwar positiv korreliert sind, aber dennoch unterschiedlich verstanden werden. Auf einer zweiten hierarchischen Ebene sind breitere Fähigkeiten angeordnet, die den primären kognitiven Fähigkeiten von Thurstone und von Cattell und Horn ähneln. Auf der höchsten (dritten) hierarchischen Ebene wird ein allgemeiner Faktor der Intelligenz erfasst. Die Hierarchie mit allen drei Ebenen beschreibt dann die Gesamtstruktur der Intelligenz. In diesem Dreischichtmodell von Intelligenz ähneln die ersten beiden Ebenen der von Cattell und Horn vertretenen Theorie der fluiden und kristallisierten Intelligenz. Diese strukturelle Anordnung besitzt Vorteile für die Beziehung zu anderen theoretischen Konzeptionen. Carrolls Theorie wird von vielen Forschern als die beste Darstellung der Struktur der Intelligenz betrachtet.

Die Theorie von von Cattell, Horn und Carroll

Die offensichtliche Ähnlichkeit zwischen dem Cattell-Horn-Modell (Cattell, 1971; Horn & Noll, 1997) und dem Carroll-Modell (1993, 1997) führte dazu, dass Forscher im März 1996

angeregt wurden, beide Modelle zur heutigen *CHC-Theorie*, die die Anfangsbuchstaben ihrer Begründer Cattell, Horn und Carroll im Namen führt, zu integrieren. Sie wurde erstmals von Flanagan, McGrew und Ortiz publiziert (2000). Obwohl Horn und Carroll die einheitliche Terminologie des Cattell-Horn-Carroll-Modells befürworteten, bestehen immer noch Unsicherheiten über die Organisation einiger Faktoren auf unterschiedlichen Ebenen. Cattell und Horn ordneten beispielsweise quantitative Fähigkeiten separat als Gq , neben Gf und Gc an, während Carroll quantitative Fähigkeiten am ehesten unter Gf subsumiert sah.

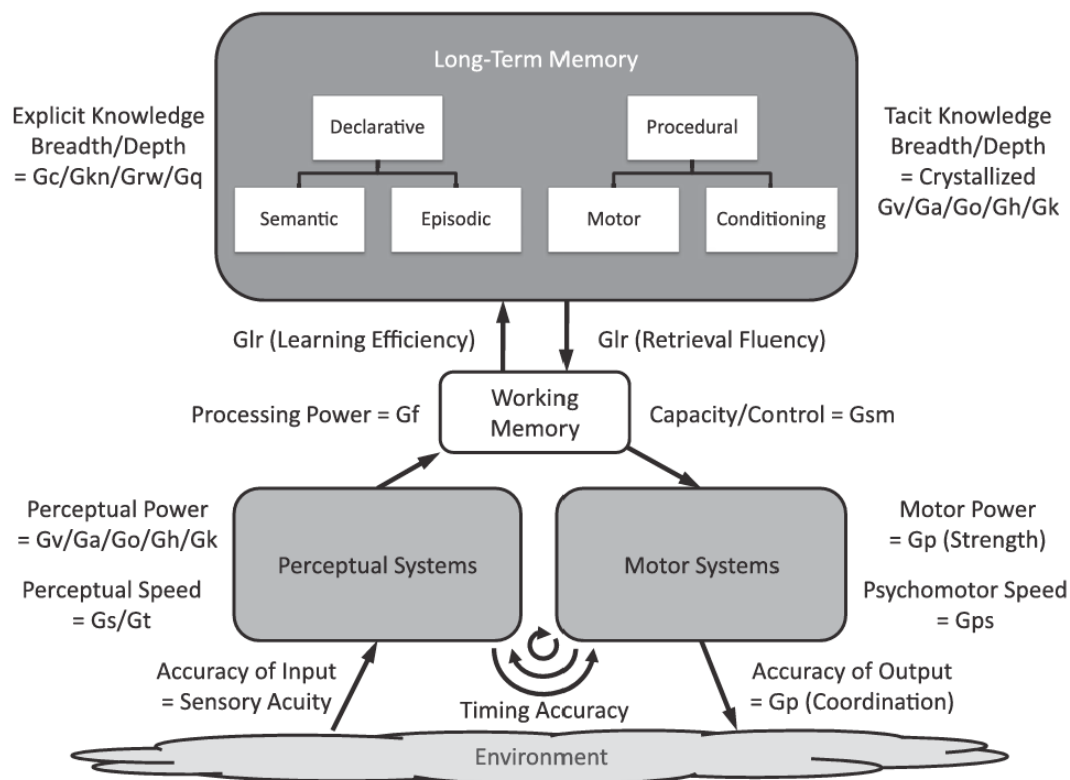


Abbildung 3 CHC abilities as parameters of information processing (aus Schneider & McGrew, 2012).

Das CHC-Modell ist vielfach theoretische Grundlage für aktuelle Studien, die den Zusammenhang zwischen schulischen (Wissens-)Leistungsmaßen und kognitiven Fähigkeiten in das Zentrum ihrer Forschungen stellen, so auch für die Studien von Kaufman (2012). Für Kaufmans Studie erscheint das CHC-Modell als theoretisches Modell geeignet, da zehn Fähigkeiten identifiziert werden, wie beispielsweise kristallisiertes Wissen (Gc), fluide Denkfähigkeiten (Gf), Kurzzeitgedächtnis (Gsm) und Prozessgeschwindigkeit (Gs). Von den

zehn Fähigkeiten lassen sich acht als kognitive und zwei als akademische Leistungsmaße identifizieren (Lesen und Schreiben), darunter auch quantitatives Wissen. Ferner ist die CHC-Theorie wichtigste Grundlage für nahezu alle individuellen kognitiven Fähigkeitstests (Kaufman et al., 2009).

Erweiterungen der CHC Theorie durch Wissenserwerbskomponenten

Obwohl die CHC-Theorie weitreichende Bedeutung erlangt hat, finden sich jedoch Kritiker, die der Ansicht sind, dass sie den Einfluss diverser Aspekte, wie zum Beispiel die Arbeitsspeicherkapazität oder Wissenserwerbskomponenten nicht genügend berücksichtigt. An erster Stelle sind hier die Verfechter der Informationstheorie zu nennen, auf die im Folgenden ein genauerer Blick geworfen wird. Im Kontext der Informationsverarbeitung sieht die Arbeitsspeichertheorie die Arbeitsspeicherkapazität als einen weiteren Aspekt der kognitiven Funktion an (Baddeley, 1986, 2000). Der Arbeitsspeicher erlaubt einem Individuum, Informationen kurzzeitig im Gedächtnis zu behalten und gleichzeitig andere mentale Operationen zu bewältigen. Baddeley konnte belegen, dass Aufgaben, die mehr Arbeitsspeicherressourcen benötigen, mit höherer Intelligenz von Personen im Allgemeinen korrelieren. Aufgrund der Ergebnisse wird diskutiert, dass sich die Arbeitsspeicherkapazität besser zur Erklärung fluider Intelligenz eignet als die mentale Geschwindigkeit (Süß et al., 1996).

Borkowski und Peck (1986) wiederum schreiben den Trainingsanteilen einen bedeutenden Einfluss auf die Gedächtnisleistungen zu. Die Autoren plädieren für die Unterscheidung in spezifische Intelligenzfaktoren. Sie gehen dabei ähnlich wie Cattell vor, der eine Unterscheidung in fluide und kristallisierte Intelligenzanteile vornimmt. Auf der übergeordneten Exekutivebene wird das erworbene Basiswissen, also auch das Bildungswissen, eines Individuums repräsentiert. Ferner sind dort das Regelwissen, das Strategiewissen und auch metakognitive Kontrollprozesse angeordnet. Diese wiederum sind auch bei der Problemlösung erforderlich und stehen ebenfalls in Wechselwirkung mit der Gedächtnisleistung. Sie werden von den Autoren als trainierbar angesehen.

Sternbergs triarchische Theorie der Intelligenz (1986) ist ein Beispiel für eine integrative Theorie, die Ansätze der Informationsverarbeitung mit psychometrischen Ansätzen zu verbinden sucht. Drei verschiedene Arten von informationsverarbeitenden Komponenten werden in der triarchischen Theorie abgebildet. Zum einen sind das metakognitive Komponenten, die verantwortlich für die Planung, die Überwachung und die Bewertung von Leistungen sind. Ferner finden sich in der Theorie Prozesse niedrigerer Ordnung, die sich mit

der Aufgabenausführung befassen. Diese können das Kodieren von Texten sein, das Vergleichen der Wortpaare und Abrufen von Speicherinformationen über Testelemente. Die Wissenserwerbskomponenten sind Lernprozesse, das Behalten und die Integration neuer Informationen mit alten Erfahrungen. In Sternbergs Komponenten-Subtheorie („componential-subtheory of intelligence“) wurden dann erstmals Wissenserwerbskomponenten als ausschlaggebend für die Intelligenzausprägung näher betrachtet (Sanchez & Sternberg, 1991). Ihr Einfluss im Kontext der Intelligenzentwicklung wurde innerhalb der Begabungsforschung erstmals von Sternberg (1986) und Sanchez und Sternberg (1991) theoretisch diskutiert. Sternberg spricht hier insbesondere elementare Informationsverarbeitungsprozesse an und unterscheidet Metakomponenten höherer Ordnung, die in der Problemerkennung und Entscheidungsfindung und zur Herangehensweise an Problemstellungen (Aufmerksamkeit, Überwachung und Kontrolle des Problemlösestandes und Evaluation) bedeutsam sind. Weiter unterscheidet er Performancekomponenten, die den Prozess der Problemlösung begleiten, wie der Gewinn neuer Informationen, das Verknüpfen und Vergleichen mit bereits vorhandenem Wissen, welches somit dazu dient, Regeln zu erkennen (deduktives Denken) und Regeln zu finden (induktives Denken). Als dritte Subkomponente nennt er die Transferkomponenten. Sie werden auch als Wissenserwerbskomponenten bezeichnet und dienen der Generalisierung und damit der Übertragung des erworbenen Wissens zu Zwecken der Verallgemeinerung bei der Ausführung der Problemlösung. Hierbei werden mittels selektiver Informationskodierung relevante von irrelevanten Informationen getrennt. Im Anschluss erfolgt die selektive Informationskombination, in der dann die relevanten neuen Informationen zu einem stimmigen Ganzen zusammengefügt werden (Perleth, 1997).

In diesem Zusammenhang kommt dem erworbenen Wissen, dem Vorwissen und – in der Erweiterung – auch dem Expertenwissen eine große Bedeutung zu. Ein kurzer Exkurs soll dieser Bedeutung, die im Zusammenhang mit den vier Forschungsfragen der vorliegenden Arbeit wieder aufgegriffen werden wird, Rechnung tragen.

Die Bedeutung des Wissens im Zusammenhang mit kognitiven Fähigkeiten bzw. in der Auseinandersetzung mit dem Begriff der Intelligenz wurde bereits im Rahmen der Gf-Gc-Theorie von Cattell (1971) betont, wonach kristallisiertes Wissen infolge des breiten Wissenserwerbs im kulturellen Kontext entsteht und darunter auch Experten- und Domänenwissen verstanden werden kann. Es entstanden weitere Modelle, die Ähnlichkeit mit der Investitionstheorie von Cattell aufweisen. Die bekannteste Definition stammt von Anderson

(2007), der deklaratives Wissen und prozedurales Wissen unterscheidet. Prozedurales Wissen bezieht sich dabei auf die Art und Weise, wie kognitive Fähigkeiten und Operationen tatsächlich ausgeführt werden. Deklaratives Wissen bezieht sich auf Fakten. In der pädagogischen Psychologie kommt eine weitere Unterscheidung hinzu. Mit konzeptuellem Wissen ist die Vernetzung von Wissenselementen gemeint, deren Wichtigkeit nicht deutlich genug hervorgehoben werden kann. So spielen Wissens- und Wissenserwerbsprozesse eine prominente Rolle sowohl in der Lernpsychologie als auch in der Expertisetheorie, die beide den Lern- und Trainingseffekten eine enorme Bedeutung beimessen. Beispielsweise ist man in der Problemlöseforschung an den kognitiven Informationsverarbeitungsprozessen und an den Wissensarten interessiert (Gruber, 2008). Im Gegensatz dazu ist die Strukturforschung eher an den Ergebnissen dieser Informationsverarbeitungsprozesse interessiert. Unterschiede zwischen den Personen werden dann mit unterschiedlichen kognitiven Fähigkeiten erklärt. Das Vorwissen als besondere Form des Wissens hat eine Bedeutung für das Problemlösen, denn es kann bestimmte Prozesse der Informationsverarbeitung entlasten (Renkl, 1996a). Vorwissen wird als dynamische Verfügbarkeit von Wissen bei speziellen Lernaufgaben definiert. Es strukturiert und enthält deklaratives, prozedurales und konditionales Wissen (Renkl, 1996a), kann darüber hinaus explizit oder implizit sein und begriffliche sowie metakognitive Wissenskomponenten enthalten. Unter speziellen Fertigkeiten ist im Zusammenhang mit kognitivem Wissenserwerb durch Lernvorgänge vor allem bereichsspezifisches Vorwissen zu verstehen. Mit der Erfahrung im Anwenden dieses Wissens kann auch neues Fachwissen schneller eingeordnet und mit bestehenden Strukturen verknüpft werden: „Offensichtlich spielt das schon vorhandene Wissen eine entscheidende Rolle für den Erwerb neuen Wissens.“ (Spada & Wichmann, 1996).

Auch in den Untersuchungen von Weinert und Helmke (1997) und Schneider und Stefanek (2004) zeigten sich größere Zusammenhänge zwischen Vorwissen und Leistung als zwischen Intelligenz und Leistung. Bereits wenn Informationen selektiert werden müssen, kann die Aufmerksamkeit direkt auf lösungsrelevante Aspekte gelenkt werden. Die Encodierung kann sogleich mit dem Vorwissen in den Zusammenhang gebracht werden. Zudem können größere Sinneinheiten gespeichert werden, da sie als Routinewissen die Ressourcen des Arbeitsgedächtnisses entlasten. Dessen Kapazität wird für weitere Informationsverarbeitungsprozesse freigehalten. Ebenso kann die Informationsspeicherung effizienter an bestehende Schemata angeknüpft werden, sodass die Vernetzung sogleich sinnvoll erfolgt und nicht später nochmals reflektiert und gegebenenfalls umstrukturiert werden muss. Das Wissen selbst erleichtert auch den effizienten Abruf relevanter Informationen aus

dem Langzeitgedächtnis. Zudem kann ein assoziatives Wissensnetzwerk die neuen Informationen zugleich an verschiedene interagierende Wissens Elemente fügen, die vielfältige Herangehensweisen ermöglichen. Zuvor noch nicht „gedachte“ Lösungswege können eher konstruiert werden. Der bereits in anderen Situationen erprobte Wissensschatz erleichtert die Anwendungsbedingungen für Problemlöseschemata. Er ermöglicht ein routiniertes Anwenden der Prozeduren, die andernfalls erst schrittweise erprobt und geübt werden müssen. Die freigehaltene Arbeitsspeicherkapazität kann zudem für metakognitive Prozesse der Überwachung eingesetzt werden (Mayer, 1992).

Die theoretische Nähe zur Investitionstheorie (Cattell) wird hier deutlich. Menschen mit höheren kognitiven Fähigkeiten fällt also der Umgang mit neuen Problemen leichter. Ihre kognitiven Fähigkeiten ermöglichen ihnen den Zugang zu einer größeren Wissensbasis und halten Arbeitsspeicherressourcen frei. Sie können sich neues Wissen zügig und effizient aneignen, welches erfolgreich an bereits bestehende Wissens Elemente angeknüpft wird, so dass in kürzerer Zeit wiederum neues Wissen vermehrt werden kann. Leistungsunterschiede resultieren somit aus den Unterschieden in den kognitiven Fähigkeiten von Personen als auch aus dem den Personen zur Verfügung stehenden Wissen. So stellen Renkl und Stern (1994) interessante Ergebnisse zum Einfluss des Vorwissens auf das Lösen von einfachen und komplexen mathematischen Textaufgaben vor. Die Daten gewannen Renkl und Stern (1994) aus der SCHOLASTIK-Studie (Schulorganisierte Lernangebote und Sozialisation von Talenten, Interessen und Kompetenzen; Weinert & Helmke, 1997). In diese Studie wurden 57 Schulklassen der Region München mit insgesamt 711 Schülern einbezogen. Sie führten einen kognitiven Fähigkeitstest CFT (Weiß & Osterland, 1979) in der deutschen Version durch. Die Autoren untersuchten 568 Grundschulkinder hinsichtlich ihrer Problemlösefähigkeiten. Ihre Studie hatte zum Ziel, die kognitiven Determinanten beim Lesen der Aufgaben zu bestimmen und ihren Beitrag zur Erklärung von Leistungsunterschieden zu ermitteln. Die Ergebnisse zeigen zunächst höhere Korrelationen zwischen komplexen Textaufgaben mit der Intelligenz als zwischen einfachen Textaufgaben und der Intelligenz. Es zeigt sich, dass fluide Intelligenz einen höheren Einfluss auf die Lösungskompetenz von mathematischen Textaufgaben hat als das Vorwissen. Darüber hinaus zeigen sich höhere Effekte, je höher das Vorwissen und die Intelligenz ausgeprägt sind. Damit wirken sich höheres Vorwissen und Intelligenz deutlich günstiger auf die Lösungsqualität der Aufgaben aus, dennoch zeigt sich die stärkste Wirkung durch einen spezifischen Vorwissenseffekt. Renkl und Stern (1994) konnten zum einen den Einfluss fluider Intelligenz und zum anderen den Einfluss von Vorwissen deutlich machen. Nur ein geringer Teil konnte durch unterschiedliche

Unterrichtsmerkmale in den 33 Klassen erklärt werden. Renkl und Stern (1994) verstehen daher unter fluider Intelligenz die Flexibilität im Umgang mit allgemeinen Prozess- und Kontrollstrategien, die beim Lösung neuer Aufgaben erforderlich ist. Schneider (1992) konnte belegen, dass das Vorwissen einen größeren Beitrag zur Aufklärung der Schulleistungen leistet als die Intelligenz. Vorwissen spielt demnach zum einen dann eine Rolle, wenn Übungseffekte erzielt werden, die vornehmlich intelligenzunabhängig sind. Faktoren der Intelligenz dienen jedoch der Kompensation fehlenden Vorwissens. Zudem spielt auch ein kumulativer Effekt von Vorwissen und Intelligenz bei der Lösung eine Rolle, der dazu führt, dass intelligenter Personen zuvor mehr Lerngelegenheiten aufgesucht haben und diese effizienter erinnern und in neuen Situationen erfolgreicher anwenden können. Diese Kumulation ähnelt dem sogenannten „Matthäus-Effekt“, der erstmals durch Stanovich (1986) bekannt gemacht wurde. Der Matthäus-Effekt bezieht sich auf ein biblisches Gleichnis, in dem gesagt wird: „Nehmt ihm nun das Talent und gebt es dem, der die zehn Talente hat. Denn wer (viel) hat, der soll noch mehr empfangen, daß er die Fülle habe; doch wer nur wenig hat, dem soll sogar das wenige genommen werden.“ (Matthäus 25, Vers 29 in Albrecht, 1926) bezieht Stanovich sich mit diesem Begriff auf die Beziehung zwischen kognitiven Fähigkeiten und schulischen Leistungen. Laut Stanovich erzielen Personen mit guten kognitiven Voraussetzungen im Vergleich zu den Personen, die weniger gute kognitive Voraussetzungen haben, im Laufe der Zeit immer weitere Zuwächse an kognitiven Fähigkeiten. Ursächlich dafür sind ihr schnelleres und effektiveres Lernvermögen, ihre überlegene Erinnerungsfähigkeit oder auch die an bessere Wahrnehmung gebundene logische Denkfähigkeit. Der Zugriff auf relevante Informationen aus dem Gedächtnis gelingt ebenfalls effizienter und macht Ressourcen frei, die weiteren Wissenszuwachs ermöglichen. Je höher das Niveau vormaliger kognitiver Fähigkeiten, umso eher lässt sich neu erworbenes Wissen effizient mit bereits bestehenden Wissensdomänen verknüpfen und in später benötigten relevanten Situationen zügig zur Verfügung stellen. Die Effizienz dabei bewirkt wiederum freierwerdende kognitive Ressourcen, die im interindividuellen Vergleich zügiger und effektiver für weitere kognitive Anforderungen genutzt werden können. Im Gegensatz dazu müssen andere Personen sich relevante Informationen erneut aneignen, sie haben zudem einen weniger effizienten Zugriff auf Gedächtnisinhalte oder auch eine geringere Speicherkapazität des Gedächtnisses, sodass sich in der Folge ein zunehmend größerer Abstand zwischen den Personengruppen abzeichnet. Daraus ergibt sich ein Vorteil gegenüber weniger intelligenten Personen. So können kognitiv fähigere Schüler zum einen durch eine hohe Auffassungsgabe schneller lernen, mehr Vorwissen aufbauen, und auf dieses tief verstandene, vernetzte

und flexibel nutzbare Wissen effektiv zugreifen und durch das effektivere Anwenden von Problemlösestrategien kumulatives Wissen aufbauen (Weinert, 1997).

Auch Renkl (1996) betont die Bedeutung des Vorwissens mit dem von Piaget beschriebenen Prinzip der Akkommodation und Assimilation. Er sieht das Vorwissen als starken Prädiktor aufgrund der festgestellten korrelativen Beziehung zwischen Vorwissen und Leistung in Höhe von .65 bis .74. Jedoch lässt sich fehlendes Vorwissen nicht vollständig durch hohe allgemeine Problemlösekompetenz kompensieren. Denn die hohe prognostische Validität von Intelligenz und Schulerfolg oder auch Berufserfolg belegten Heller (1991, 1997), Weinert und Helmke (1993), Weinert und Hany (2000) sowie Stumpf und Schneider (2010) und Perleth (1997). Dennoch habe das Vorwissen bedeutenden Einfluss auf die zu erbringende Leistung. Langfeldt (2006) belegt beispielsweise den im Vergleich zu den kognitiven Fähigkeiten zunehmenden Einfluss des Vorwissens anhand der zunehmenden Korrelation im Laufe der Lerndauer. Helmke und Schrader (2010) dagegen nehmen an, dass Schulleistungen hauptsächlich aufgrund der Vorkenntnisse vorhergesagt werden können und kognitive Fähigkeiten eine untergeordnete Rolle spielen. So wird an dieser Stelle deutlich, dass kein Konsens zu dieser Wirkbeziehung besteht.

Expertiseforscher, die sich mit Wissen in Form von Expertise beschäftigen, erweitern aber die bisherigen Vorstellungen kognitiver Fähigkeiten um die dynamische Vorstellung der Leistungsentwicklung, die abhängig ist von in Lernprozesse investierte Zeit und Trainingsintensitäten (Ericsson, Krampe & Heizmann, 1993). Expertise beinhaltet die Fähigkeit, in einem spezifischen Bereich zielgerichtet handeln zu können, Problemstellungen zu bewältigen und Lösekompetenz für Aufgaben zu besitzen sowie die Verfügbarkeit von speziellem Wissen in diesem Bereich. Die Entstehung von Expertise wird nach drei theoretischen Expertiseansätzen nach Gruber und Mandl (1996) kategorisiert: Zum einen existieren Expertisemodelle, die vorwiegend einen dispositionalen Aspekt favorisieren. Diese beruhen auf Vorstellungen von besonderen Begabungen, die Personen zum Experten werden lassen. Daneben wird der Aspekt, der vor allem umfangreiche Übung als wichtigsten Prädiktor beim Expertisegewinn betrachtet, hervorgehoben. Beide Ansätze diskutieren (Ericsson, 1985; Gruber & Mandl, 1992). Eine zweite Kategorie der Expertiseansätze bilden jene, die Experten vor allem mit genereller Kompetenz ausgestattet sehen, die über überdurchschnittliche allgemeine Problemlösestrategien verfügen, welche sie in vielfältigen Problemsituationen erfolgreich anwenden können. Andere Vorstellungen betonen vor allem bereichsspezifische spezielle Kenntnisse, die die Personen in einer bzw. in wenigen Domänen als Experten ausweisen. Eine dritte Kategorie beschäftigt sich mit dem erworbenen

Wissen und dem Gedächtnis von Experten. Hierbei dominiert vor allem der Einfluss des Vorwissens gegenüber anderen Fähigkeiten, wie zum Beispiel kognitiven Fähigkeiten oder generellen Problemlösestrategien, auch Metakognitionen. Ackerman (1989) formulierte folgende Komponenten von Expertise: Übung als eine Funktion des Lernens begünstigt das Korrelationsverhältnis zwischen Fähigkeit und Leistung. Weiterhin wirken allgemeine Fähigkeiten insbesondere zu Beginn der Leistungssituation besonders förderlich. Mit zunehmender Übung gewinnen aufgabenspezifische Faktoren an Bedeutung. Bei inkonsistenten Aufgaben besitzen dagegen allgemeine kognitive Ressourcen eine größere Bedeutung, da keine vollständige Routine entstehen kann. Diese wird eher bei konsistenten Aufgabenanforderungen erworben, sodass dann die allgemeinen Fähigkeiten weniger bedeutsam sind.

Zusammenfassend lässt sich sagen, dass die Bedeutung des Wissens, des Vorwissen und auch des Expertenwissens nicht unterschätzt werden kann. Die Verfechter der oben genannten Ansätze sehen sich in ihrer Idee bestätigt, dass ein umfassendes Theoriemodell auch diese Aspekte integrieren muss und sehen die Tatsache, dass die CHC-Theorie diese nicht integriert, als einen Schwachpunkt des Modells.

Auch über die oben genannten Aspekte hinausgehende Kritik am CHC-Modell blieb ebenfalls nicht aus. So bemerkt Schneider (1992) in seinem Artikel zur Relevanz kognitiver und nicht kognitiver Einflussfaktoren auf den Expertiseerwerb, dass es schwierig sei, spätere Spitzenleistungen über früh erfasste psychometrische Intelligenztestverfahren vorherzusagen. Vielmehr seien nicht-kognitive Faktoren wie zum Beispiel Motivation und Ausdauer eher in der Lage, spätere Spitzenleistungen vorherzusagen. Der Nutzen traditioneller Intelligenztests zur Vorhersage des Studienerfolges wird aus diesem Grund auch immer wieder in Frage gestellt, da Fachwissensprädiktoren den Studienerfolg beispielsweise geeigneter vorhersagen konnten (Mulhauser, 2001; Ackerman, 2003). Ackerman untersuchte den Zusammenhang zwischen Intelligenz und Bildungsleistungen unter den Aspekten, welchen Einfluss Persönlichkeitsfaktoren und motivationale Faktoren auf Unterschiede im Bereich des Wissens einnehmen. Infolgedessen lässt sich die auf Cattell zurückgehende Gf-Gc-Theorie insofern von Ackerman (1996) erweitern, als dass auch Expertise und Domänenwissen, die als breiter Wissenserwerb im kulturellen Kontext erworben werden, integriert werden. Diese Faktoren integrierte Ackerman in seinem PPIK-Modell. Die Abkürzung steht für die Merkmale Intelligenz als Prozess, Persönlichkeit, Interessen und Intelligenz als Wissen. Das Modell übernimmt damit auch Einflüsse des Domänenwissens und Persönlichkeitsfaktoren aus dem Cattell-Modell und weist damit Ähnlichkeiten mit Cattells Investitionstheorie (1971, 1987) auf.

Ericsson und Smith (1991) gehen sogar so weit zu sagen, dass moderierende Effekte der intellektuellen Fähigkeiten kaum vorhanden sind. So gehen Ericsson und Crutcher (1990) als Vertreter der Expertisetheorie davon aus, dass außergewöhnliche Leistungen nicht durch angeborene basale kognitive Fähigkeiten determiniert sind, sondern bereichsspezifisch erworben werden. Sie seien darüber hinaus ohne intensive Übung nicht zu erreichen, was wiederum eine hohe individuelle Anstrengungsbereitschaft und Motivation voraussetzt. Dennoch machen auch Gruber und Mandl (1996) darauf aufmerksam, dass in kürzester Zeit erworbene Fähigkeiten kaum ohne dispositionale Fähigkeiten erklärt werden können.

Schneider (1993) versuchte daraufhin, beide Ansätze zu vereinen, indem er ein „Schwellenwertmodell“ zur Erklärung vorschlägt. Nach diesem sind zunächst basale Fähigkeiten ausschlaggebend, um überhaupt Leistungen in einem Bereich erzielen zu können. Überschreiten kognitive Voraussetzungen einen überdurchschnittlichen Bereich, den sogenannten Schwellenwert, dann entscheiden nicht-kognitive Voraussetzungen wie Engagement, Ausdauer, Konzentration und Erfolgsmotivation, sowie Faktoren der Lernumwelt über den zukünftigen Expertisestatus. Schneider trägt damit den Feststellungen Rechnung, Übung allein sei nicht in der Lage, Fähigkeitsunterschiede in Gedächtnisleistungen und Verständnis zu kompensieren.

Auch Luo, Thompsen und Detterman (2003) führen weitere übungsunabhängige Faktoren an, wie den Einfluss mentaler Geschwindigkeit, die etwa 30 % der Varianz zwischen kognitiver Fähigkeit und schulischer Leistung erklären kann. Rhode und Thomsen (2007) untersuchten die Rolle des Arbeitsgedächtnisses und der räumlichen Vorstellungsfähigkeiten sowie der Bearbeitungsgeschwindigkeit und fanden wiederum Belege für die Rolle dieser spezifischen kognitiven Fähigkeiten.

Die prädiktive Gültigkeit der Gf-Gc-Theorie der Wirksamkeit spezifischer kognitiver Fähigkeiten wurde von Flanagan und McGrew (1997) belegt. Im Zusammenhang mit den Mathematikleistungen fanden Floyd, Evans und McGrew (2003) moderate Beziehungen zwischen mathematischen Leistungen und kristallisierter Intelligenz. Darüber hinaus stellten sie ebenso moderate Beziehungen zu fluider Intelligenz, dem Kurzzeitgedächtnis und dem Arbeitsspeicher fest. Insgesamt fanden sie, dass Intelligenz als Prädiktor für mathematische Leistungen gilt (McGrew, Flanagan, Keith & Vanderwood, 1997).

Neuere Theorien versuchen verstärkt mentale Prozesse, wie Planungsfähigkeit und Aufmerksamkeit zu integrieren. Daneben existieren theoretische Auffassungen, dass auch praktische Intelligenz, Kreativität, künstlerische oder musikalische Fähigkeiten erfasst wer-

den sollten. So unterscheidet Sternberg vor allem analytische, kreative und praktische Fähigkeiten und entwickelte ein entsprechendes Testverfahren (Sternberg & Powell, 1983; Sternberg, 1986).

Nachfolgend erweiterten sich die theoretischen Vorstellungen über die Bereiche der Intelligenz bis zu der Vorstellung „multipler Intelligenzen“ von Gardner (1993). Er argumentierte zunächst für acht Intelligenzen, bestehend aus räumlichem Vorstellungsvermögen sowie sprachlicher, logisch-mathematischer, musikalischer, körperlich-kinästhetischer, naturalistischer, interpersonaler sowie intrapersonaler Intelligenz. Seine Identifikation der Intelligenzen basiert auf empirischen Nachweisen (Gardner 1993, 2006), wobei einschränkend gesagt werden muss, dass diese Intelligenzen schwierig zu messen sind, wie etwa das Engagement von Kindern, und daher die Ergebnisse als nicht belastbar vorausgesetzt werden können.

Die Theorie von Jäger, Süß und Beauducel

Im Laufe der Forschung wurde den oben genannten Faktoren wie Motivation und Interesse sowie anderen Persönlichkeitsmerkmalen, die zur Erklärung von Leistungsexzellenz geeignet sind, weitere Bedeutung zur Prognose von schulischem und akademischem Erfolg beigemessen. Es wurden neuere mehrdimensionale Faktorenmodelle, wie das differenzierte Berliner Intelligenzstrukturmodell von Jäger, Süß und Beauducel (1997) konzeptionalisiert, das Cattell verpflichtet ist. Es bezeichnet auf der übergeordneten Ebene die allgemeine Intelligenz mit g , doch berücksichtigt es im Rahmen der Betrachtung von Problemlöseleistungen bereichsspezifische kognitive Leistungen und auch Operationmerkmale, wie Bearbeitungsgeschwindigkeit, Gedächtnis, Einfallsreichtum und Verarbeitungskapazität, die eine bedeutende Rolle hinsichtlich der Qualität der Leistungserbringung spielen. Es werden figural-bildhafte Fähigkeiten, verbale Fähigkeiten und numerische Fähigkeiten unterschieden. Das Modell besteht aus zwei Dimensionen mit insgesamt sieben Hauptkomponenten. Dabei werden auf der einen Seite vier Komponenten als "Operation" und auf der anderen Seite drei Komponenten als „Inhalte“ abgebildet. Mit den vier unter „Operation“ zusammengefassten Komponenten werden vier Kategorien von Fähigkeiten verstanden. Diese sind die Bearbeitungsgeschwindigkeit (B), das Gedächtnis (G), der Einfallsreichtum (E) und die Verarbeitungskapazität (K). Auf der anderen Dimension werden unter „Inhalte“ die drei Komponenten und Fähigkeitsbündel figural-bildhafte Fähigkeiten (F), verbale Fähigkeiten (V) und numerische Fähigkeiten (N) unterschieden. Damit ordnen die Autoren jeder Intelligenztestaufgabe eine 4×3 -Matrix zu, die dann die entsprechende Fähigkeitskombination

zur Lösung erfordert. Die Autoren verstehen ihre bimodale Matrix als Repräsentation multifaktorieller Intelligenzanteile, die zur Lösung der entsprechenden Aufgabenbereiche angesprochen werden. Das Berliner Intelligenzstrukturmodell wurde anhand umfangreicher Stichproben entwickelt und empirisch national und international validiert. Es gilt somit als gut fundierte Abbildung der menschlichen Intelligenz.

Die Theorie von Perleth und Sierwald

Ein weiteres Beispiel für die Konzeption mehrdimensionaler Modelle ist das von Heller (1992) und Perleth und Heller (1994) entwickelte Münchner (Hoch-)Begabungsmodell. In diesem Modell werden, wie in der Abbildung dargestellt, verschiedene Begabungsfaktoren angenommen, die dann in Leistung gezeigt werden können, wenn auf der einen Seite die Begabungsfaktoren mit den nicht-kognitiven Persönlichkeitsmerkmalen, wie Stressbewältigung, Leistungsmotivation, Arbeitsverhalten, (Prüfungs-)angst und Kontrollüberzeugungen günstig interagieren. Auf der anderen Seite moderieren Umweltmerkmale, wie familiäre Faktoren, Instruktionsqualität, Klassenklima und kritische Lebensereignisse die Leistungsmerkmale.

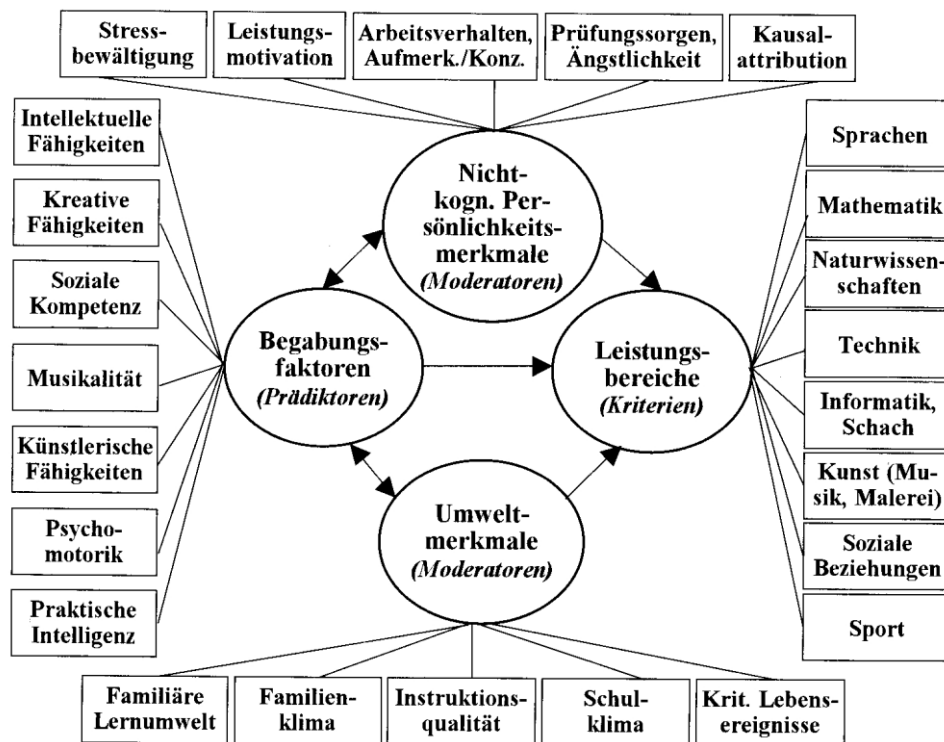


Abbildung 4 Münchner Hochbegabungsmodell (aus Perleth, Joswig & Hoese, 2012).

Im Rahmen einer von Perleth durchgeführten Follow-up-Untersuchung zur Münchner Hochbegabungsstudie im Jahr 1997 wurde das Münchner (Hoch-)Begabungsmodell (MMG) um Ansätze der Expertiseforschung, die explizit das Vorwissen einbeziehen, erweitert. So berücksichtigt das erweiterte Münchner Dynamische Begabungs-Leistungsmodell (Perleth, 1997) die Bereichsspezifität außergewöhnlicher Leistungen, lange Übungszeiten und den gesamten Verlauf der Entwicklung außergewöhnlicher Leistungen, vollzieht die Trennung von Wissensaspekten und allgemeinen Kompetenzen und integriert die Interaktion von Anlage-Umwelt-Beziehungen (Perleth, 1997, 2001). Das Münchner Dynamische Begabungs-Leistungsmodell, das auch als Munich Dynamic Ability-Achievement-Model (MDAAM) bezeichnet wird, bildet somit eine Brücke zwischen dem prospektiven Diagnose-Prognose-Ansatz der Hochbegabungsforschung einerseits und dem retrospektiven Experten-Novizen-Paradigma der Expertiseforschung andererseits (Perleth, 1997, 2001). Als angeboren geltende Informationsverarbeitungsmerkmale, wie Aufmerksamkeit, Aufmerksamkeitssteuerung, Habituation, Gedächtniseffizienz, Aktivationsniveau und Aspekte der Wahrnehmung oder Motorik, stellen den Rahmen der individuellen Begabungsentwicklung dar (Perleth, 2001).

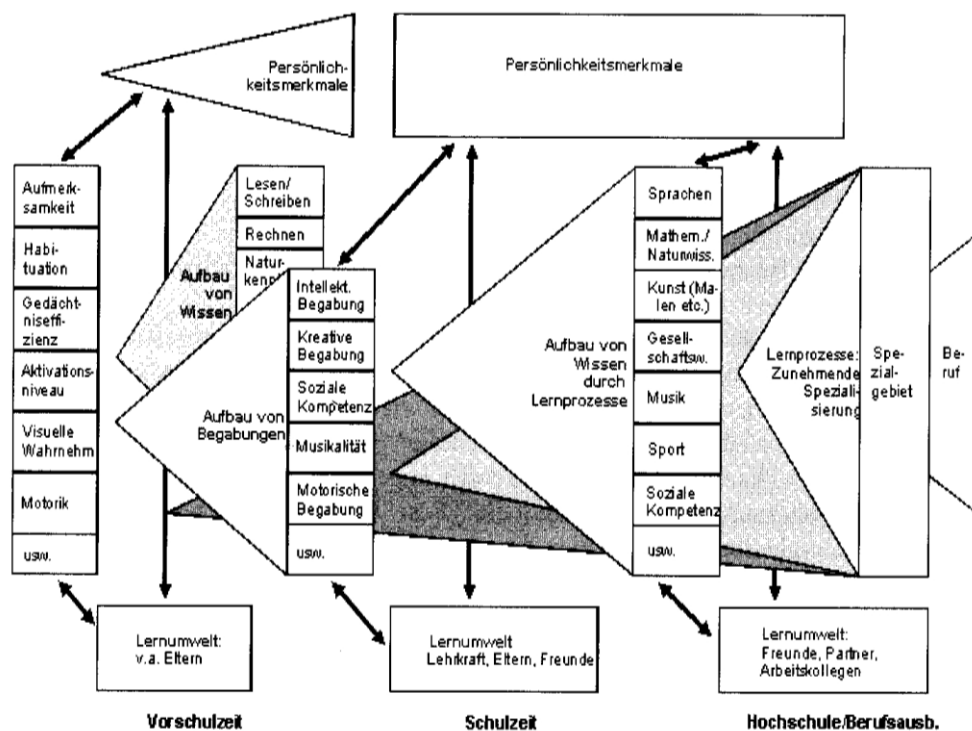


Abbildung 5 Das Münchner dynamische Begabungs-Leistungs-Modell (MDAAM) (aus Perleth, Joswig & Hoese, 2012).

Die Konzeption von Perleth und Sierwald (1996) auf der Grundlage des oben genannten Begabungsmodells von Perleth und Heller (1994) erfolgt in Anlehnung an das Berliner Intelligenzstrukturmodell, argumentiert jedoch für ein gleichsames Vorhandensein eines Generalfaktors im Sinne allgemeiner Intelligenz sowie die Existenz der verschiedenen unabhängigen Faktoren. Perleth und Sierwald berücksichtigen dabei auch vorwissensabhängige (reasoningabhängige) Dimensionen. Diese sind zum einen nach dem Aufrufen von Regeln bzw. Wissen (Deduktion) und dem Finden von Regeln (Induktion) zu unterscheiden.

Das Modell unterscheidet drei bzw. vier Phasen der Leistungs- bzw. Expertiseentwicklung, die an die Phasen der vorschulischen, schulischen und beruflichen bzw. universitären Entwicklung angelehnt sind (Perleth, 2001). Entwicklungsprozesse werden in der folgenden Abbildung durch Dreiecke dargestellt. Bereits in der frühen Kindheit, wenn elterliche Sozialisationseinflüsse besonders bedeutsam sind, können Anregungen richtungsweisend für die weitere Begabungsentwicklung sein. Aufgrund der angeborenen Begabungsmerkmale und der Aktivitäten bilden sich Begabungsschwerpunkte heraus. Gleichzeitig werden wichtige Grundlagen für die Persönlichkeitsentwicklung gelegt (Dreieck links oben in der Abbildung), die ebenso wie Leistungsmotivation, Durchhaltevermögen usw. für die Begabungsentwicklung bedeutsam sind (Perleth, 2001).

Vielfältige Aktivitäten der Kinder in diesem Alter bewirken den Aufbau von Fertigkeiten und Wissen bis zu elementaren Rechenkenntnissen. Eine gute Passung zwischen Anlagefaktoren und (elterlich gestalteter) Lernumwelt wirkt sich besonders förderlich für die Begabungsentwicklung dieses Alters aus. Mit zunehmendem Alter der Kinder gewinnen Pädagogen an Einfluss auf die Begabungsentwicklung der Kinder. Mit ungünstigen familiären Gestaltungsmöglichkeiten wird die individuelle Förderung der Begabungsschwerpunkte umso wichtiger (Perleth, 2001). Zur Begabungsentwicklung kommt der Aufbau von Wissen und Fertigkeiten in den unterschiedlichen Bereichen hinzu (helles Dreieck mittig in der Abbildung). Dieses Wissen stellt die entscheidende Grundlage für die weitere Leistungsentwicklung dar. Begabung ist wichtig, um Wissen aufzubauen. Andererseits sind erst umfangreiches und gut organisiertes Wissen sowie gut entwickelte Fertigkeiten die Basis für Leistungsentfaltung. Gute Leistungen im Erwachsenenalter setzen Wissen und Begabungen voraus. Im Jugendalter gewinnen Freunde der Heranwachsenden zunehmenden Einfluss auf die Begabungs- und Leistungsentwicklung. Meist werden gleichinteressierte und gleichbegabte Freunde gesucht, mit denen die Begabungen weiterentwickelt werden. Im Erwachsenenalter ist die berufliche Spezialisierung (helles Dreieck rechts in der Abbildung) die zentrale Aufgabe der Leistungsentwicklung. Diese gelingt umso besser und schneller,

je günstiger die Begabungskonstellationen im Hinblick auf das jeweilige Fachgebiet ausgebildet werden und je solider die in der Schulzeit erworbenen Fertigkeiten, Kenntnisse und Wissensbestände sind. Bei manchen setzt die Spezialisierung zum Teil schon in der Schul- oder Vorschulzeit ein (mittel- und dunkelgraue Dreiecke in der Abbildung). So beginnen beispielsweise Schachspieler, Sportler oder Musiker sehr früh mit speziellem Training, um im Erwachsenenalter Spitzenleistungen zu erzielen (Heller und Perleth, 2007). Das von Perleth (1997) entwickelte Modell ist für diese Arbeit wegen seiner bereichsspezifischen Konzeption kognitiver Fähigkeiten und der Berücksichtigung des Einflusses von Vorwissenselementen im Sinne Cattells bedeutsam, da die Konzeption des dieser Arbeit zugrundeliegenden kognitiven Testverfahrens darauf basiert.

2.1.1.2 Testverfahren zur Messung kognitiver Fähigkeiten

Nachdem nun mittels der Darstellung der verschiedenen grundlegenden Theorien eine Annäherung an den Begriff kognitive Fähigkeiten in seinen vielfältigen Ausprägungen gelungen ist, steht die Auseinandersetzung mit den im Laufe der Forschung entwickelten Testverfahren, die zur Überprüfung der Theorien entworfen wurden, im Mittelpunkt. Nach einem kurzen historischen Abriss wird besonderes Augenmerk auf die Vorzüge, aber auch auf die Mängel der einzelnen Testverfahren gerichtet. Damit soll die Grundlage für die im empirischen Teil der Arbeit verwendeten Tests und ihre Aussagekraft gelegt werden. In einem ersten Schritt werden zunächst die standardisierten Gütekriterien der klassischen Testtheorie vorgestellt.

Der Begriff Test (lat.: testimonium) kann wörtlich mit Zeugnis, Prüfung oder Beweis übersetzt werden. Tests im Sinne von Tüchtigkeitsproben oder auch Eignungsverfahren waren bereits in vielen frühen Gesellschaftsformen etabliert (Heller, 1973). Mit einem Intelligenztest werden individuelle Messwerte erfasst, die darüber Auskunft geben, inwieweit eine Abweichung vom Mittelwert der Vergleichsgruppe besteht (Krapp & Weidenmann, 2001). Eine solche Bestimmung mittels Testverfahren dient als Erkenntnismethode in der Human- und Sozialwissenschaft (Eid, Gollwitzer & Schmitt, 2013). Im Rahmen psychologischer Diagnostik werden Intelligenztests als Verfahren zur quantitativen bzw. qualitativen Bestimmung von kognitiven Personenmerkmalen eingesetzt (Krapp & Weidenmann, 2001).

Klassische Testtheorie

Die Qualität dieser Testverfahren kann mit Hilfe der Gütekriterien der klassischen Testtheorie bewertet werden. Nach Lienert werden diese in Haupt- und Nebengütekriterien unterschieden. Letztere sind die Ökonomie, Nützlichkeit, Normierung und Vergleichbarkeit der empirischen Messungen (Lienert & Raatz, 1994).

Die Hauptgütekriterien der klassischen Testtheorie sind zunächst die Objektivität, mittels derer die Unabhängigkeit der Resultate der Messung vom Untersucher unabhängig sind. Ein Testverfahren ist objektiv, wenn im optimalen Fall jeder Untersucher dasselbe Ergebnis feststellt (Ingenkamp & Lissmann, 2008). Die Ansprüche an die Objektivität richten sich an die Bedingungen der Durchführung, die unabhängig vom Untersucher sind, wenn Testsituation, Testmaterial und Testumgebung den gleichen Bedingungen unterliegen (Durchführungsobjektivität). Weiter ist die Auswertungsobjektivität erreicht, wenn die Messergebnisse unabhängig von der Person des Auswerter das selbe Ergebnis liefern. Mit Hilfe standardisierter Auswertungsverfahren über Schablonen, Kategoriensysteme und erwartete Antworten kann dieses Ziel erreicht werden. Die Interpretationsobjektivität ist erreicht, wenn kein sogenannter Ermessensspielraum vorliegt, sondern die Leistung anhand von Normierungstabellen mit der Vergleichsgruppe eingestuft werden kann. Die Objektivität ist eine notwendige Voraussetzung für die Vergleichbarkeit der Ergebnisse und damit für die Validität einer Untersuchung (Bortz, 2005).

Die Reliabilität als zweites Hauptgütekriterium gibt die Genauigkeit eines Messverfahrens an. Es wird erwartet, dass ein Test exakt die gemessenen Merkmale differenziert, unabhängig davon, ob er dieses Merkmal auch zu messen beansprucht (Bortz, 2005). Obwohl die Reliabilität mathematisch exakt definiert ist, kann sie empirisch nicht präzise bestimmt werden, da der wahre Wert eines Personenmerkmals nicht vollkommen exakt ermittelt werden kann. Daher beruht die Schätzung auf der Berechnung eines Vertrauensintervalls, das die Abweichung des gemessenen Wertes vom wahren Wert in die Interpretation einfließen lässt. Mittels der Reliabilitätsschätzung wird die Abschätzung der Fehlervarianz eines Tests bestimmt (Eid, Gollwitzer & Schmitt, 2013).

Die Validität gilt als drittes Hauptgütekriterium der klassischen Testtheorie und gibt die Gültigkeit eines Testverfahrens an, also ob das Testverfahren das misst, was es messen soll (Bortz, 2005). So soll ein Aufmerksamkeitstest dieses Konstrukt auch erfassen und nicht ein anderes (z. B. die Reaktionsgeschwindigkeit). Für Aussagen über die Validität eines

Testverfahrens ist das Vorhandensein von Kriterien erforderlich. In Anlehnung an die Empfehlungen der American Psychological Association APA (1974) werden drei Arten von Gültigkeiten unterschieden. Die Unterscheidung erfolgt dabei in Inhaltsvalidität, mittels derer überprüft wird, ob das Messverfahren das zu messende Merkmal angemessen repräsentiert. Inhaltsvalidität ist gegeben, wenn ein Test eine repräsentative Stichprobe (Teilmenge) aus einer Gesamtaufgabenmenge (Aufgabenuniversum) darstellt, die zur Messung eines Personenmerkmals von einem Expertengremium definiert und als geeignet zur Messung eines bestimmten Merkmals bezeichnet wurde (Bortz, 2005). Die Kriteriumsvalidität spiegelt mittels regressionsanalytischer Verfahren die Beziehung zu einem Kriterium wider. Die Kriteriumsvalidität wird durch den Grad der Übereinstimmung zwischen der Leistung im Test und einem oder mehreren Validitätskriterien bestimmt. Kriteriumsvalidität liegt vor, wenn die Korrelation beispielsweise eines Intelligenztests zu einem anderen Intelligenztest hoch ist. Diese zeigt an, dass beide Tests in sehr hohem Grade das Gleiche messen. Der Validitätskoeffizient eines Tests wird durch die Korrelation des Tests und äußerem Kriterium ermittelt. Wenn die Testresultate kongruent sind, so besteht eine Übereinstimmungsvalidität. Eine Vorhersagevalidität liegt vor, wenn die Korrelation der Testwerte mit Testwerten, die zu einem späteren Zeitpunkt erhoben worden sind, hoch ist. Die Vorhersagevalidität besitzt eine besondere Bedeutung bei der Prognose von zukünftigen Leistungen. Die Konstruktvalidität dagegen ist die Information darüber, ob ein Testverfahren das zugrundeliegende Konstrukt, das gemessen werden soll, tatsächlich erfasst. Die Ergebnisse eines Testverfahrens werden in den Zusammenhang mit anderen Tests gestellt, die eine Beurteilung bereits existierender gültiger Indikatoren des zu messenden Konstrukts erlauben (Bortz, 2005). Ein Konstrukt stellt dabei eine theoretische Variable dar, die nicht vollständig operational definierbar ist. Es ist unmöglich, ein Konstrukt wie Intelligenz durch eine exakt umrissene Menge von Verhaltensweisen oder Merkmalen zu beschreiben. Es ist stattdessen innerhalb einer psychologischen Theorie definiert und muss erschlossen werden. Das bedeutet, dass insofern aus dem Konstrukt Hypothesen abgeleitet werden, deren empirische Betätigung durch die Höhe der ermittelten Korrelationen erfolgt. Die Faktorenanalyse als Verfahren zur Feststellung der Konstruktvalidität dient der Feststellung, welche Testaufgaben „zusammengehören“, also das Gleiche erfassen. Die Konstruktvalidität beschreibt die Beziehung zwischen dem Testergebnis und dem theoretischen Konstrukt durch das Ausmaß, in dem von einem Testergebnis auf die Ausprägung eines Konstrukts geschlossen werden kann. Eine niedrige Konstruktvalidität bedeutet den Einfluss von Messfehlern. Zum einen kann ein anderes als das intendierte Konstrukt gemessen werden und zum anderen ist

es möglich, dass die zugrundeliegende Theorie nicht den realen Verhältnissen entspricht (Bortz, 2005).

Konzeptualisierung von Intelligenztestverfahren

Testverfahren zur Messung kognitiver Fähigkeiten sind nach den oben kurz beschriebenen testtheoretischen Grundsätzen konstruiert und halten den standardisierten Gütekriterien der Objektivität, Validität und Reliabilität, den Überlegungen der klassischen Testtheorie, stand. Die Einschätzung, also die Messung der Intelligenz, ist zusammen mit ihrer Konzeptualisierung seit über einem Jahrhundert ein Schwerpunkt der Forschung. Schullaufbahnentscheidungen werden nicht selten aufgrund von kognitiven Fähigkeitstestverfahren getroffen. Beispielsweise wird der Zugang zum Gymnasium häufig über eine Kombination aus kognitiven Fähigkeitstests und anderen zulassungsrelevanten Auswahlverfahren gesteuert. Auch Zugangsmöglichkeiten zu bestimmten Zusatzangeboten, wie zu Sommerakademien, Enrichment-Angeboten der Begabtenförderung und zu bestimmten Schulformen oder auch Klassen, die speziell hochbegabte Kinder fördern, basieren häufig auf kognitiven Fähigkeitstests. Andererseits sind Intelligenztestverfahren aber auch valide Diagnoseinstrumente bei Verdacht auf Teilleistungsschwächen oder Lernstörungen und können eine Intervention oder eine Zuweisung in eine Förderschule legitimieren. Theoretischer Hintergrund ist dabei die Vorstellung, dass das Niveau der kognitiven Fähigkeiten das Niveau der Schulleistungen bestimme oder eben auch begrenze (Grube, 2008). Da jedoch – wie dargelegt – keine Einigkeit über die Definition von Intelligenz herrscht, besteht auch keine Einigkeit darüber, wie sie zu messen ist. Darüber hinaus werden der Begriff Intelligenztest und der Ausdruck "IQ-Test" austauschbar verwendet, obwohl sie nicht gleichwertig sind (Urbina, 2011). Einige Forscher wie Spearman und Thurstone definieren Intelligenz als das, was Intelligenztests messen (Sternberg & Detterman, 1986). Inzwischen existieren etwa 3500 Testverfahren zur Messung kognitiver Fähigkeiten (Mackintosh, 2011). Obwohl sie gleichermaßen wertvoll sind, unterscheiden sich Menschen, so berichtet Urbina (2011), in vielerlei Hinsicht voneinander. Kognitive Fähigkeiten bzw. Intelligenz ist einer dieser Unterschiede zwischen den Menschen. Intelligenztests sind daher seit Spearmans ersten Versuchen (1904), kognitive Fähigkeiten zu messen, unabdingbare Werkzeuge zur Erläuterung individueller kognitiver Unterschiede. Ein Problem jedoch war damals und ist bis heute die eindeutige statistische Messung.

Francis Galton entwickelte die ersten Tests mit objektiven Techniken, um Intelligenz zu messen. Er behauptete nie, dass seine Tests direkt die Intelligenz messen, da sie

eher auf physikalische Maßnahmen als auf höhere mentale Prozesse abzielen (Wasserman & Tulsky, 2005; Mackintosh, 2011). Auch Cattell versuchte, geistige Fähigkeiten systematisch zu messen. Sein Test umfasste zusätzlich zu Galtons sensorischen und motorischen Tests solche der höheren kognitiven Aufgaben. Alfred Binet befasste sich mit den Einschränkungen von Galton und Cattells Bewertungsmaßstäben, entwickelte 1905 zusammen mit Théodore Simon die Binet-Simon-Skala und begründete so die Psychometrie. Im Gegensatz zum Galton- und Cattell-Maß für motorische und sensorische Maßnahmen ist Binets Skala eine Bewertung komplexer geistiger Fähigkeiten (Wasserman & Tulsky, 2005). Die Tests umfassten die Messung von Sprache, Lernen und Gedächtnis, Urteilsvermögen und Problemlösungsfähigkeit sowie auditive und visuelle Verarbeitung und sollten zunächst dazu dienen, Pariser Schüler mit Lernschwierigkeiten zu identifizieren, die pädagogische Hilfe benötigten (Urbina, 2011). Binet und Simon zogen 1905 (in Urbina, 2011) zunächst das kindliche Alter als Kriterium zur Messung heran. Sie stellten fest, dass achtjährige Kinder mehr Items einer Testbatterie korrekt beantworten konnten als jüngere Kinder. Das führte zu der Weiterentwicklung der Tests, um das Alter der Kinder als Norm zwar festzuhalten, aber die Differenzen zwischen den Kindern zu bestimmen. Henry Goddard führte 1910 die Binet-Simon-Skala in Amerika ein. Dort wurde sie 1916 von Lewis Terman an der Stanford University adaptiert und nunmehr als Stanford-Binet-Intelligence-Scale bezeichnet (Wasserman & Tulsky, 2005). Hier fiel auf, dass einige jüngere Kinder ebenso viele korrekte Antworten geben konnten wie ältere Kinder. Diese erste Skala wurde daher ebenfalls mehrmals überarbeitet und für ältere Altersgruppen erweitert.

Die Raven-Matrizentests wiederum wurden 1936 von John C. Raven als sprachfreie Papier-und-Bleistift-Tests in Form von Multiple-Choice-Tests konzipiert und sollten es ermöglichen, den erfolgreichen Umgang von Personen mit abstrakten Begriffen zu messen. Vorteil dieser Tests ist ihre ökonomische Effektivität. Auch sie ermöglichen die zeitgleiche Erfassung von Fähigkeiten im Rahmen großer Gruppen von Personen. Diesem Vorteil steht jedoch auch ein entscheidender Nachteil gegenüber, denn in dieser Konzeption können nur nonverbale Fähigkeiten im Sinne einer allgemeinen Intelligenz erfasst werden, nicht aber spezifische Fähigkeiten. Um diesen Limitationen entgegenzuwirken, begann die Entwicklung von Tests, die geeignet waren, spezifische Fähigkeiten zu messen. Integriert werden sollten vor allem Wissensanteile, die im Bildungssystem erworben werden. Die Testverfahren wurden dahingehend weiterentwickelt, dass sie die verschiedenen Komponenten der Intelligenz abbilden, also ein Stärken- und Schwächenprofil der getesteten Person erstellen

können. Edward Lee Thorndike entwickelte aufgrund der angenommenen Bereichsspezifität als einer der Ersten Intelligenztests, die verbale, quantitative und nonverbale Fähigkeiten erfassen (1978). Unter den folgenden Testverfahren, die dieser Multidimensionalität Rechnung tragen, erreichten die von David Wechsler konzipierten Wechsler-Tests, die erstmals 1944 veröffentlicht wurden (Wechsler, 1944; Mackintosh, 2011), bedeutenden Einfluss. Sie erfassen vier Bereiche kognitiver Fähigkeiten, wie sprachliches Verständnis, wahrnehmungsgebundenes logisches Denkvermögen, Arbeitsgedächtnis und Prozessgeschwindigkeit. Entscheidender Unterschied zu den bis dahin etablierten Stanford-Binet-Tests war die Erfassung der Intelligenz erwachsener Personen. Diese Innovation Wechslers wurde durch die Ermittlung der Differenzen zwischen leichten und schweren Items möglich. Wechslers Konzept der Abweichungsintelligenz war eine Weiterentwicklung der Konzeption der Stanford-Binet-Skalen.

Ab Ende der 1970er Jahre fand eine Entwicklung in Richtung hierarchischer Modelle innerhalb der psychometrischen Tradition statt. Auf der Basis von Cattells Theorie entstand der Versuch, Tests zu entwickeln, die sowohl einen allgemeinen, fluiden Faktor der Intelligenz erfassen als auch spezifische Fähigkeiten. Die überwiegende Mehrheit der Testverfahren zur Messung fluider Fähigkeiten beruht auf Intelligenztestaufgaben und verwendet nonverbale, relativ kulturfreie Stimuli, sie erfordern die Integration von verbalem und nonverbalem Denken. Anhand von sprachfreiem, figuralen Material wird die allgemeine Fähigkeit getestet, in neuartigen Situationen Denkprobleme zu erfassen, Beziehungen herzustellen, Regeln zu erkennen sowie Merkmale zu identifizieren und rasch wahrzunehmen (Weiß & Osterland, 1979). Da die fluiden Intelligenzanteile sich weitgehend unabhängig von soziokulturellen Faktoren entwickeln, können sie mit einem entsprechenden Testverfahren, das induktives und deduktives Denken kulturübergreifend erfasst, gemessen werden. Cattell (1940) bezeichnete diesen Test als Culture Fair Test (CFT). In der Konsequenz kann sich für eine Person also ein Profilbild ergeben, in der ihre verschiedenen Fähigkeiten abgebildet werden, so zum Beispiel ein hohes Niveau verbaler Fähigkeiten und der Fähigkeit zu schlussfolgern, während diese Person zugleich Schwierigkeiten im räumlichen Vorstellungsvermögen haben kann. Es folgten Forschungsbemühungen, die den Einfluss von Informationsverarbeitungsprozessen (Verarbeitungsgeschwindigkeit und Arbeitsgedächtnis) sowie die neurobiologischen Einflüsse oder auch die Aspekte der Metakognition, der Planung, des Wissens, der beruflichen Fähigkeiten und des Fachwissens untersuchten (Ackerman, 1996; Sternberg, 1985; Deary et al., 2007).

Andere Testverfahren beruhen auf der Annahme eines multidimensionalen Verständnisses kognitiver Prozesse und werden als Fähigkeitstests verwendet, um spezifische Fragestellungen hinsichtlich des Fähigkeitsprofils zu beantworten. So werden Intelligenz- bzw. kognitive Fähigkeitstests optimal eingesetzt, um Antworten auf gezielte Fragestellungen zu finden (Perleth, Joswig & Hoese, 2012). Ein auf dieser Grundlage entwickelter Intelligenztest erfasst demnach ein ganzes Profil intellektueller Fähigkeiten. Auch der kognitive Fähigkeitstest (KFT 4 bis 13+) von Heller, Gaedike und Weinländer (1985) basiert auf dieser Modellkonzeption. In ihrer Konzeption des kognitiven Fähigkeitstests stützen Perleth und Sierwald (1996) sich vor allem auf Thurstones unabhängige Faktoren, gruppieren diese aber unter Akzeptanz eines auf die Bereiche wirkenden g-Faktors zu den Fähigkeitsbündeln verbale, nonverbale und quantitative Fähigkeiten (Perleth & Sierwald, 1996; Perleth, 1997).

Zusammenfassend lässt sich festhalten, dass der Einsatz von Fähigkeitstests von einigen Forschern kritisiert wird, da kein voll umfängliches Testverfahren zur Erfassung sämtlicher Fähigkeiten existiert (McGrew, Flanagan, Keith, & Vanderwood, 1997). Andererseits sind Testverfahren, die auf der Erfassung eines allgemeinen Intelligenzfaktors g basieren, zwar zumeist ökonomisch und effektiv einzusetzen, besitzen wiederum den Nachteil, dass sie sehr unspezifische Ergebnisse liefern und nur eingeschränkt gezielte Stärken- und Schwächenprofile aufzeigen können.

Aktuelle Testentwicklungsforschung konzentriert sich auf die Konzeptionalisierung von faktoranalytisch orientierter Fähigkeitserfassung, die sowohl allgemeine als auch spezifische Fähigkeiten auf unterschiedlichen Ebenen erfasst. McGrew, Flanagan und Ortiz (2000), McGrew, Flanagan, Keith und Vanderwood (1997) sowie Alfonso, Flanagan und Radwan (2005) begründen die Notwendigkeit, spezifische Fähigkeiten im Sinne *Gf-Gc* zu erfassen und die hierarchische Beziehung zu einem g-Faktor zu berücksichtigen. Sie entwickelten Strukturgleichungsmodelle und validierten im Einklang mit Carrolls 3-Stratum-Modell der Intelligenz eine neue Generation von Intelligenztestbatterien. Sie belegen, dass spezifische Fähigkeiten für die Entwicklung spezifischer Lese- und Mathematikfähigkeiten nötig sind. Damit wird über das bisherige Verständnis im Sinne eines g-Faktors und den Leistungskonstrukten hinausgegangen.

Ziel von Testverfahren

Tests zur Einschätzung kognitiver Fähigkeiten werden vor allem zur Vorhersage von allgemeinem Bildungserfolg eingesetzt. Unabhängig davon, ob ein bereichsspezifisches Verständnis kognitiver Fähigkeiten angenommen wird oder ein allgemeiner Intelligenzfaktor g

als konzeptionelle Grundlage vorausgesetzt wird, es gelten kognitive Fähigkeiten in der synonymen Verwendung als ein starker Prädiktor späteren akademischen Erfolgs (Gottfredson, 1998), denn Gottfredson stellte fest, dass Intelligenz, gemessen mit einem standardisierten Testverfahren, der einzige effektive Prädiktor für individuelle Leistungen in Schule und Beruf ist. Diese Aussage wird von den meisten psychologischen Forschern anerkannt, jedoch nicht von allen, wie bereits eingangs dieser Arbeit betont wurde. Die Wirkung von Intelligenz auf künftige akademische Leistung wurde beispielsweise auch in einer Crossed-Lagged-Panel-Studie von Watkins, Lei und Canivez (2007) untersucht. Sie testeten 289 Schülerinnen und Schüler mit einem Testintervall von 2,8 Jahren und berichteten, dass der Intelligenzquotient eine starke Wirkung auf akademische Leistungen hatte. Willis, Dumont und Kaufman stellten fest, dass die meisten verbalen Subtests in Intelligenzskalen primär als die Messung der kristallisierten Intelligenz klassifiziert werden (2011). Da kristallisierte Intelligenz erworbenes Wissen darstellt (Cattell, 1963; Horn, 1991), sind Tests, die dieses abfragen, vergleichbar mit Leistungstests (Cattell & Horn, 1978). Allgemeine Intelligenztest- und Leistungskorrelation werden übereinstimmend als positiv berichtet (Herrnstein & Murray, 1994). Nach Gustafsson und Undheim (1996) beträgt die Korrelation typischerweise etwa .50. Mackintosh (2011) berichtet Korrelationen in Höhe von .40 bis .70 zwischen Intelligenz und Schulleistungen, während Gottfredson (2005) Korrelationen zwischen Intelligenz und standardisierten Schulleistungstests um rund .60 berichtet. Deary et al. (2007) berichten über Zusammenhänge bis .81. Deshalb werden die standardisierten Intelligenztests hauptsächlich als Prognoseinstrument eingesetzt, um die zukünftige akademische Leistung vorherzusagen (Deary, Strand, Smith & Fernandes, 2007) oder um die Eignung für Begabtenförderungen oder den Bedarf an individuellen Interventionsmaßnahmen festzustellen (Gustafsson & Undheim, 1996; Jensen, 1998). Intelligenztests werden weiterhin zur Feststellung des Lernfortschritts im Rahmen von Längsschnittuntersuchungen eingesetzt. Sie sind für Pädagogen einfach zu verwalten, zeitwirksam, zuverlässig und replizierbar (Urban, 2011). Das wird jedoch auch durchaus kritisch betrachtet. Dabei wird weniger ihre allgemeine Validität angezweifelt als ihr unangemessener Einsatz, wenn beispielsweise Zugangsmöglichkeiten zu bestimmten Bildungsinstitutionen gesteuert werden sollen (Rost, 2005).

Zusammenfassend lässt sich sagen, dass Tests zur Messung der Intelligenz zu den zuverlässigsten und gültigsten psychologischen Tests und Bewertungen zählen (Gottfredson, 1997). Die meisten der zeitgenössischen Tests sind direkte Nachkommen, Modifikationen oder Reaktionen der ersten Testverfahren von Binet und Simon (Urban, 2011).

Sie wurden jedoch aufgrund von Forschungserkenntnissen entsprechend der sich differenzierenden Forschungstraditionen weiterentwickelt. Somit unterscheiden sich die heutigen Intelligenztestverfahren nach ihrer konzeptionellen Grundlage. Obwohl die Intelligenztests in der Literatur häufig kritisiert werden, sind die meisten Kritikpunkte mit der unangemessenen Verwendung dieser Tests und der Fehlinterpretation der Ergebnisse, aber nicht mit den Tests selbst verbunden (Gottfredson, 1997). Standardisierte Maßnahmen der Intelligenz prognostizieren heute viele wichtige pädagogische, berufliche, wirtschaftliche und soziale Ergebnisse (Gottfredson, 1997; Brody, 1997). Aber es muss trotz des belegten Zusammenhangs zwischen Intelligenz und Bildungserfolg auf die Ergebnisse der sozialwissenschaftlichen Forschung hingewiesen werden. Sie besagen, dass der Bildungserfolg vornehmlich von primären und sekundären Bildungsherkunftseffekten oder auch von institutionell bedingter Chancenungerechtigkeit abhängt (Boudon, 1974; Berger & Kahlert, 2008). Gerade bei Fragen zur weiteren Schullaufbahn oder der Eignung für bestimmte weiterführende Bildungsinstitutionen ist daher die Erfassung spezifischer Fähigkeiten nützlicher (Perleth, 1997).

2.1.2 Mathematische Schulleistungen

Nach der oben erfolgten komplexen Annäherung an den Begriff kognitive Fähigkeiten, seine Verwendung im Rahmen dieser Arbeit und die zugehörigen Theorien und Intelligenztests soll nun in einem nächsten Schritt die ähnliche komplexe Definition für den Begriff mathematische Schulleistungen mit der Skizzierung entsprechender Modelle erarbeitet werden. Diesem Schritt kommt ein hoher Stellenwert zu, denn mathematische Schulleistungen werden als eine Hauptdeterminante schulischen Erfolges betrachtet und besitzen für den schulischen Kontext daher einen mindestens ebenso großen Einfluss wie kognitive Fähigkeiten. Außerdem sind mathematische Fähigkeiten Basis naturwissenschaftlicher Schulfächer und werden vielfach in Auswahlverfahren für Berufs- und Studienzulassungen abverlangt.

In diesem Zusammenhang ist es zur Eingrenzung des Terminus notwendig, zunächst kurz auf den Begriff Schulleistungen im Allgemeinen und schließlich mathematische Schulleistungen im ganz Speziellen einzugehen.

Ganz allgemein wird Leistung als zielgerichtete Handlung bezeichnet, die ergebnisorientiert ist. Dabei ist die Beurteilung, inwieweit das Ziel erreicht wird, das Endresultat eines Beurteilungsprozesses, bei dem ein vorgegebenes Qualitätskriterium vorausgesetzt

wird (Jäger, 2007). Akademische Leistung wird von Jäger als der Grad der Fähigkeiten verstanden, wie vorgegebene Tatsachen und Konzepte sowohl quantitativ als auch qualitativ durchgeführt werden. Im Zusammenhang mit der akademischen Leistung wird zunächst als der Grad der Fähigkeit, wie vorgegebene Tatsachen und Konzepte sowohl quantitativ als auch qualitativ durchgeführt werden, betrachtet. Da Leistung ein breites und multidimensionales Konstrukt ist, kann die Definition oder das Verständnis des Konstrukts unterschiedliche Dimensionen betonen (Jäger, 2007).

Das Konzept der pädagogischen Leistung, das im Rahmen dieser Arbeit besonders interessiert, wird speziell in der Sozialwissenschaft verwendet und bezieht sich insbesondere auf die Beherrschung von Wissen und Fertigkeiten. Es kann auch als Leistung auf einem bestimmten Gebiet definiert werden. Aus schulpädagogischer Perspektive wird Leistung daher als die Fähigkeit zur Durchführung von Anforderungen in akademischen Bereichen (wie Lesen und Mathematik) konzipiert (Schipolowski et al., 2014). Die Autoren ziehen eine enge Verbindung zwischen kristallisierter Intelligenz und akademischer Leistung. So wird akademische Intelligenz als Fähigkeit, aus Erfahrung zu lernen oder Wissen zu erwerben, verstanden, während akademische Leistung als das Gelernte oder das Wissen betrachtet werden kann. Der Zusammenhang zwischen Intelligenz und Leistung stand auch in den Forschungsbemühungen von Jensen (1998) im Zentrum. Er stellte aufgrund seiner Literaturrecherchen fest, dass der allgemeine Intelligenzfaktor g mit schulischen Leistungen, der Reaktionszeit, der Arbeitsleistung, dem Berufsstatus, dem Einkommen und der Kreativität korreliert. Zahlreiche andere Studien haben die Beziehung zwischen Intelligenz und Leistung ebenfalls untersucht. Obwohl Methoden und Datenerfassungstechniken variieren, zeigen die meisten, dass Intelligenz und Leistung stark korreliert sind. Infolgedessen gilt das Verhältnis zwischen allgemeiner Intelligenz und akademischem Erfolg als gut etabliert (Jensen, 1998). Intelligenz und Leistung sind zwei verschiedene, aber miteinander verknüpfte Konstrukte. Diese enge Verzahnung regte zahlreiche Forscher an, Unterschiede in Bildungsleistungen zu erklären und zu einem besseren Verständnis der Beziehung zwischen Intelligenz und Leistung beizutragen.

Auf der Grundlage dieser und ähnlicher empirischer Ergebnisse wurden und werden weitreichende bildungspolitische Investitionen initiiert, um die Lernerfolge von Schülerinnen und Schülern positiv zu beeinflussen. Schulleistungsforschung konzentriert sich daher hauptsächlich auf Wissenserwerbsprozesse in der Schule und bemüht sich zum Beispiel um die möglichst optimale Ausgestaltung didaktischer Planungen. So weisen die Schulgesetze explizit auf die Erforderlichkeit des Erwerbs von Fachwissen in den Unterrichtsfächern hin

und formulieren Bildungsstandards, welche die Vergleichbarkeit von Lernfortschritten gewährleisten sollen (Ministerium für Bildung, Wissenschaft und Kultur Mecklenburg-Vorpommern (2010, 2015). Dabei orientieren sich diese Bildungsstandards zunächst an dem Kompetenzansatz, der nach den Ebenen Handlungskompetenz, Methodenkompetenz, Sachkompetenz, Personale Kompetenz und Sozialkompetenz unterschieden wird. Unter Kompetenzen werden die bei Individuen verfügbaren kognitiven Fähigkeiten und Fertigkeiten verstanden, um bestimmte Probleme erfolgreich zu lösen (und die damit vorhandene motivationale, volitionale und soziale Bereitschaft und Fähigkeit (Weinert, 2001). Unter Handlungskompetenz wird unter anderem das Zusammenwirken von kognitiven und sozialen Fähigkeiten und Fertigkeiten verstanden. Hinsichtlich der Methodenkompetenz steht der Erwerb von Lernstrategien, Arbeitstechniken und Techniken der Informationsbeschaffung im Vordergrund. Insbesondere die Entwicklung von Sachkompetenz, die als zweiter Bildungsstandard formuliert wurde, wird über die Entwicklung systematischer Fähigkeiten, Fertigkeiten und Kenntnisse in der Auseinandersetzung mit Inhalten, Aufgaben und Problemen im schulischen Kontext aufgebaut. Dabei gilt es zunächst, den Umgang mit mathematischen Problemen zu erlernen und im Unterricht Begriffe und Zusammenhänge aus verschiedenen Inhaltsbereichen kennenzulernen (Weinert, 2001). Mathematische Methodenkompetenz wiederum orientiert sich vor allem daran, in der Konfrontation mit mathematischen Problemen Erkenntnisprozesse zu forcieren, die zu der korrekten Auswahl einer Lösungsstrategie führen. Die Auseinandersetzung als problemorientierte Herangehensweise wird im Mathematikunterricht besonders betont. Hinter der Modellierung von Kompetenzdimensionen steht die Erwartung, „[...] Wissen und Können in der Schule so zu vermitteln, dass keine ‚trägen‘ und isolierten Kenntnisse und Fähigkeiten entstehen, sondern anwendungsfähiges Wissen und ganzheitliches Können“, erzielt werden (Klieme & Rakoczy, 2008). Für das Fach Mathematik sollen beispielsweise allgemeine Kompetenzen der Problemlösung, des Modellierens, Darstellens, Kommunizierens und Argumentierens erworben werden. Zudem steht der Erwerb inhaltlicher Kompetenzen im Umgang mit Zahlen, Raum und Form, Messen, Daten, Zufall und funktionalen Zusammenhänge im Zentrum des Mathematikunterrichts (Blum, 2006). Aus didaktischer Perspektive sind mathematisches Verständnis und mathematische Problemlösefähigkeit Ziele, die in Lehrplänen genannt werden. Auch Köller und Baumert (2002) sehen mathematische Schulleistung vor allem als Ergebnis schulischer Lernprozesse. Stern (1997) betont ebenfalls die Bedeutung der Schule für den mathematischen Kompetenzerwerb. Die Prävention von Lernschwierigkeiten besitzt bei dem Erwerb

spezifischer mathematischer Vorläuferfertigkeiten eine große Bedeutung für schulische mathematische Leistungen (Krajewski, 2006). Sie betont die professionelle Vermittlung bereits im Vorschulalter und belegt deren Effekte. Für den schulischen Kontext besitzen mathematische Schulleistungen insofern eine besondere Bedeutung, als dass sie als eine Hauptdeterminante des Schulerfolgs betrachtet werden (Heller, 1991, 1984, 1997; Weinert & Helmke, 1997). Mathematische Fähigkeiten sind Basis naturwissenschaftlicher Schulfächer und werden vielfach in Auswahlverfahren für Berufs- und Studienzulassungen abverlangt. Sie sind eng verknüpft mit dem Begriff mathematische Kompetenzen, der weiter gefasst ist. Während mathematische Kompetenz bedeutet, mathematisches Wissen im Lebensalltag erfolgreich anzuwenden, werden mathematische Fähigkeiten häufig nur auf Rechenfertigkeiten reduziert. Deshalb erfolgt in dieser Arbeit zunächst eine Klärung des Begriffs mathematische Kompetenz, der eine Eingrenzung auf mathematische Leistungen folgt, die in der Schule Gegenstand von Leistungsmessungen sind.

Unter mathematischer Kompetenz (Literacy) wird im Rahmen der Definition der internationalen Schulleistungsuntersuchung PISA (Programme for International Student Assessment) zunächst ganz allgemein die Fähigkeit zum Gebrauch mathematischen Wissens und Könnens auf unterschiedlichen Ebenen verstanden. Dabei ist der Einsatz von mathematischen Standardverfahren im Sinne „mathematischen Denkens“ gemeint (DPK – Deutsches PISA-Konsortium, 2001). Darüber hinaus sind Kenntnisse der verschiedenen Inhaltsbereiche aus der Mathematik und deren Anwendung eingeschlossen. Bei diesen Bereichen handelt es sich nach der PISA-Definition um Veränderungsmessungen, Wachstum, Raum und Form, die Wahrscheinlichkeitsrechnung, sowie den Umgang mit gegenseitigen Abhängigkeiten von Größen, den Umgang mit Zufallsdaten und den Umgang mit quantitativen Größen. Erst in zweiter Linie werden die Bereiche Algebra, Arithmetik und Geometrie genannt, um einer unerwünschten Verengung auf das Verständnis von Mathematik als Rechenfertigkeit vorzubeugen.

Weiter werden unter mathematischen Kompetenzen allgemeine Fähigkeiten wie mathematische Sprechweisen, das Modellieren von Situationen und ein „verständiger Einsatz“ von Kenntnissen zum Lösen mathematischer Problemsituationen verstanden. Explizit wird erwähnt, dass Querverbindungen hergestellt werden müssen sowie die Fähigkeit zur Verallgemeinerung und zum Verstehen von Zusammenhängen erforderlich ist. Auf der nächsten Ebene werden im Rahmen mathematischer Kompetenzen auch persönliche und berufliche Bildungsziele formuliert. Die mathematische Modellierungsfähigkeit wird als beson-

ders wichtig erachtet, denn sie besitzt Relevanz für das Verständnis mathematischer Probleme im Lebenszusammenhang (Baumert, Brunner, Lüdtke & Trautwein, 2009). Daher findet sie sich als zentrales Ziel in den Zielformulierungen der deutschen Rahmenpläne, die als bundesweit geltende Bildungsstandards die Ziele und Inhalte des Unterrichts vorgeben und auf deren Grundlage die schulinternen Lehrpläne erstellt werden.

Die Aufgaben in der Schule werden sowohl nach Stoffgebieten als auch nach der Art der mathematischen Herangehensweise unterschieden. Ebenso folgt die Konzeption mathematischer Schulleistungstests, wie beispielsweise von Krajewski, Küspert und Schneider (2002), Krajewski, Liehm und Schneider (2004) sowie Roick, Gölitze und Hasselhorn (2004, 2006) der Unterscheidung nach stoffgebietsspezifischen Fähigkeiten. Die Stoffgebiete werden in die Bereiche Arithmetik, Geometrie, Algebra, Stochastik eingeteilt (Baumert, 2001; Baumert & Köller, 2002). Dabei besitzen die Ziele Problemlösen und Erwerb eines generellen mathematischen Verständnisses Priorität (Sekretariat der Ständigen Konferenz der Kultusminister, 2004). Auch Köller und Baumert (2002) sprechen sich für die stoffspezifische Unterscheidung aus, da die Unterscheidung in die Sachgebiete Zahlen und Zahlenverständnis, Geometrie, Algebra, Darstellung und Analyse von Daten und Wahrscheinlichkeitsrechnung, Messen und Maßeinheiten, Proportionalität und Faktoren analytisch unterscheidbar ist. Anhand der internationalen Vergleichsstudie von mathematisch-naturwissenschaftlichem Unterricht (TIMSS, siehe Baumert & Lehmann, 1997) konnte Köller (2008) ebenfalls Aufgabenkategorien aus sechs mathematischen Stoffgebieten unterscheiden: Geometrie, Analysis, Stochastik, Algebra sowie die Anwendung und Lösung von Routineverfahren und komplexen Verfahren. Die stoffgebietsspezifische Differenzierung von mathematischen Leistungen ist auch international etabliert (National Council of Teachers of Mathematics, 2000 in Brunner, 2006).

2.1.2.1 Entwicklung mathematischer Fähigkeiten

Zu fragen ist nun, wie Kinder in die Lage versetzt werden, den oben beschriebenen Ansprüchen gerecht zu werden. Daher soll zunächst ein Blick auf die Entwicklung mathematischer Fähigkeiten vor dem Hintergrund der relevanten theoretischen Modelle geworfen werden, bevor die Entwicklung mathematischer Schulleistungen im Grundschulalter in den Fokus rückt.

Der Umgang mit Zahlen gilt in unserer Gesellschaft als kulturelles Werkzeug, welches Voraussetzung für aktive Teilhabe ist. Beispielsweise werden numerische Ausdrücke

wie die Anzahl recht anschaulich genutzt, andere numerische Ausdrücke wirken relativ abstrakt wie Prozentangaben oder die Berechnung von Zinsen. Die Fähigkeiten zu ihrer Nutzung müssen erst erworben werden, einschließlich des Lesens und Schreibens von Zahlen, dem Zählen, den vier Grundrechenarten, der Suche von Seiten in einem Buch beispielsweise oder der Auswahl eines Fernsehkanals. Diese Fähigkeiten sind mehr oder weniger komplex und Kinder müssen erst lernen, beispielsweise die Ziffer 4 nicht ausschließlich als Zählzahl zu begreifen, sondern auch zu verstehen, dass sie ein Autokennzeichen oder eine Telefonnummer bezeichnen und dadurch wie ein Eigenname oder ein Etikett verstanden werden kann und zusammen mit anderen Ziffern vorkommen kann. Darüber hinaus muss begriffen werden, dass Zahlen auch häufig in einer festen Reihenfolge präsentiert werden. Nach Piaget (1969, 1972a, 1976) ist es, für den Erwerb arithmetischer Kompetenzen notwendig zu verstehen, dass die Arithmetik logischen Prinzipien folgt und damit ein Teil der Logik ist. Damit ist auch die Fähigkeit zur Abstraktion von Wahrnehmungseigenschaften gemeint, worunter das Verständnis zu betrachten ist, dass alle Einzelheiten von Gegenständen in Farbe, Form und Größe recht unterschiedlich sein und dennoch dieselbe Zahl besetzen können und damit als ein Set von Bestandteilen zu betrachten sind. Für den Erwerb arithmetischer Fähigkeiten sind wiederum Arbeitsspeicherfähigkeiten, räumliche Kognition (Baddeley, 1986, 2000, 2006) und auch linguistische Fähigkeiten (Bloom, 1971) erforderlich. Die Korrelationen zwischen diesen Fähigkeiten und Arithmetiktests sind gut etabliert (Bull & Johnston, 1997). Nach Butterworth (2005) vollzieht sich die Entwicklung arithmetischer Fähigkeiten als Teil spezifischer kognitiver Fähigkeiten in einem zunehmend anspruchsvolleren Verständnis von Numerosität und ihrer Implikationen. Der Begriff Numerosität wird als die Fähigkeit begriffen, kleine Mengen abzuschätzen. Von Geary (1995, 2000) wird diese Entwicklung auch als sekundäre mathematische Fähigkeit formuliert. Zunehmend entwickelt sich darüber hinaus eine Geschicklichkeit bei der Manipulation dieser Zahlssysteme. Die Fähigkeit zur Unterscheidung, ob eine Menge kleiner oder größer geworden ist, wird von Geary (2000) als *simple arithmetic* bezeichnet, die sich auch bereits im Säuglingsalter ausprägt. Studien mit Säuglingen im Alter von 9 Monaten zeigen, dass sie bereits ein Grundverständnis arithmetischer Probleme besitzen, z. B. ob eine Menge mehr oder weniger geworden ist. Im Zahlenraum bis 60 können Kinder im Vorschulalter agieren, wenn die Probleme in Vergleichsaufgaben eingebettet werden (Grube & Barth, 2004; Grube, 2006, 2008). Das Verständnis von Größer-Kleiner-Relationen, welches sich etwa im Alter von 18 Monaten nachweisen lässt, wird als *ordinality* bezeichnet. Auch das counting (inneres Zählen) kann bereits in diesem Alter festgestellt werden (Spelke, 2000). Um ein Verständnis

von Zahlwörtern und ihrer Bedeutung zu entwickeln, muss auch ein präzises numerisches System aufgebaut werden, was mehrere Jahre des Lernens erfordert (Langhorst, Ehlert & Fritz, 2012 in Fritz et al., 2013). Und obwohl es sehr einfach erscheint, zählen zu lernen, dauert es also etwa vier Jahre, bis Kinder diese Fertigkeit meistern. Fortgeschrittene Zählfähigkeiten besitzen Kinder etwa mit sechs Jahren (Gelman & Gallistel, 1978 in Fritz et al. 2013). Nachdem das Zählwortprinzip (Steffe et al. in Fritz et al., 2013) verstanden wurde, erweitert sich das Konzept auf eine mentale Abbildung von Zahlenlinien (Fritz et al., 2013), mit der dann einfache arithmetische Operationen durchgeführt werden können. Das Verständnis dieser mentalen Zahlenlinie ist entscheidend für den Erwerb mathematischer Kompetenzen im Grundschulalter (Fritz et al., 2013). Das Erkennen einer Kardinalzahl (das letzte Objekt in einer Reihung), z. B. wenn man verlangt, etwa bis drei zu zählen, erfolgt ebenfalls im Vorschulalter. Es gilt als erworbene Fähigkeit (Fritz et al., 2013). Diese einfachen mathematischen Fähigkeiten entwickeln sich etwa bis zum Schuleintritt, wenn dann komplexere Fähigkeiten wie Zählen und einfache Arithmetik zur Verfügung stehen. Bei Schuleintritt haben Kinder bereits eine Reihe an Schlüsselfaktoren mathematischer Kenntnisse erworben. Kinder mit soliden Vorkenntnissen haben somit gute Chancen beim Erlernen der mathematischen Kompetenzen im Schulalter (Krajewski, 2008; Krajewski et al., 2006). Als vierte Stufe des Aufbaus mathematischer Kompetenzen gilt die Entwicklung des Verständnisses des Teil-Ganzes-Konzepts (Resnick, 1983 in Fritz, 2013). Das Kompensationsprinzip mit numerischer Genauigkeit „ $5 + 3$ “ und „ $2 + 6$ “ wird im Grundschulalter verstanden.

Das von Fritz (2013) entwickelte Niveaustufenmodell der mathematischen Kompetenzentwicklung besitzt besondere Relevanz für den Bereich Arithmetik und ist empirisch validiert worden. Für arithmetische Kompetenzen ist die Erkenntnis darüber wichtig, was mathematische Operationen sind, sowie Prozesswissen, wie diese durchzuführen sind. Ein wegweisendes Entwicklungsmodell früher mathematischer Kompetenzen wurde auch von Krajewski (2008) präsentiert. In ihm werden drei Kompetenzebenen angenommen. Numerische Basisfertigkeiten werden auf Ebene I erworben, während ein unpräzises und dann präzises Anzahlkonzept auf der Ebene II folgen. Die dritte Ebene der Entwicklung gilt dann als erreicht, wenn das Verständnis von Anzahlrelationen und Anzahlzerlegung sowie das Verständnis von Mengenunterschieden gelungen sind.

Aufbauend auf den primären Fähigkeiten wird im Grundschulalter der Erwerb von Fähigkeiten auf sekundärer Ebene zum Lösen von Aufgaben der Geometrie, Algebra und Arithmetik (Geary, 2000) möglich.

Mathematische Problemlösefähigkeit

Im Kontext der Entwicklung mathematischer Fähigkeiten ist auch Problemlösefähigkeit erforderlich. Sie wird diese auch in dem den PISA-Untersuchungen zugrundeliegenden Modell als mathematisches Problemlösen bezeichnet und explizit als relevante Kompetenz erwähnt. Für das mathematische Problemlösen sind sowohl Wissen und als auch Informationsprozesse erforderlich (Mayer, 1992; DeCorte & Verschaffel, 1987; Cragg & Gilmore, 2014). Zunächst ist sprachliches, also semantisches Wissen erforderlich, um die Problemstellung zu erfassen. Für die Eingrenzung und Einordnung des Problems ist weiter schematisches Wissen über mathematische Problemarten generell notwendig (Mayer, 1992; Bull & Johnston, 1997). Die Auswahl einer geeigneten Strategie zur Lösung erfordert wiederum strategisches Wissen, das auch den Lösungsprozess überwacht und Zwischenschritte prüft (Mayer, 1992). Zum Anwenden mathematischer Operationen ist weiter prozedurales Wissen erforderlich. Mathematik erfordert eine Fülle von Fähigkeiten und daher ist es wichtig, die verschiedenen Prozesse zu untersuchen, um die Schwierigkeiten von Kindern im Umgang mit mathematischen Problemen zu verstehen.

Im Zuge dieser Problematik befassten sich Forscher mit der Rolle der kognitiven Prozesse, wie Kinder diese Repräsentation entwickeln. Aus Forschungsergebnissen ist bekannt, dass Kinder mit mathematischen Leistungsproblemen Schwierigkeiten bei der angemessenen Problemrepräsentation haben (Goldin & Shteingold, 1998). Dabei können zum einen Schwierigkeiten bei der Übersetzung des Problems mithilfe von sprachlichem und semantischem Wissen die mentale Repräsentation beeinträchtigen. Andererseits können Schwierigkeiten bei der angemessenen Problemintegration entstehen, wenn mit der mentalen Repräsentation relevante Informationen selektiert werden und mittels des schematischen Wissens schlüssig zu einer mentalen Struktur organisiert werden müssen. Die im Anschluss der Problemlösung zugeordneten Prozesse erfordern teilweise gleichzeitig ablaufende Teilprozesse, wie das Planen der Lösung und die Überwachung der Lösungsausführung. Strategisches Wissen ist wiederum erforderlich, um die geeigneten Problemlöseprozesse zu selektieren und die eigene Tätigkeit zu kontrollieren. Auch das Textverständnis muss hinreichend entwickelt sein, so Goldin und Shteingold (1998). Sie gehen davon aus, dass die extrinsischen Schwierigkeiten der inneren Repräsentation der Konzepte durch kognitive Hindernisse alsbald bei Misserfolg auch zu affektiven Hindernissen werden. Anhand der

Schwierigkeiten von Kindern, negative Zahlen zu repräsentieren, empfehlen sie, die Entwicklung dieser inneren Repräsentationen zu fördern, um kognitive Hindernisse zu überwinden. Damit gehen sie von einer interaktiven Beziehung kognitiver Fähigkeiten und der ausgleichenden Wirkung des Wissenserwerbs aus. Dass jedoch mathematische Problemlösefähigkeit von kognitiven Prozessen abhängt, stellten Agostino et al. fest (2010). Sie untersuchten den Einfluss der mentalen Aufmerksamkeit beim Lösen mathematischer Probleme. In ihrer Stichprobe waren 155 8- bis 13-jährige Kinder, denen Probleme auf zwei Schwierigkeitsstufen präsentiert wurden. Mittels Strukturgleichungsmodellen sollte gezeigt werden, inwiefern latente Aufmerksamkeitsfaktoren die Arbeitsspeicherkapazität hemmen und somit die Problemlösung beeinträchtigen. Sie fanden, dass eine geringere Arbeitsspeicherkapazität längere Aufgabenlösung erfordert. Durch die Schwierigkeit, gleichzeitige Informationen im Gedächtnis zu behalten, wird außerdem die Arbeitsspeicherkapazität überlastet und somit werden mehr Fehler produziert. Auch De Corte und Verschaffel (1987) bestätigten, dass Schülerfehler eher auf der falschen Problemdarstellung als auf falschen Rechenoperationen selbst beruhen. Hauptproblem sei demnach eine mangelnde Fähigkeit, die Art des Problems korrekt darzustellen und eine daraus folgende unangemessene Strategieauswahl.

Baddeley legte hingegen den Fokus auf die Rolle des Arbeitsgedächtnisses beim Lösen mathematischer Aufgaben, da es an einer Reihe von mathematischen Aktivitäten beteiligt ist (Bull & Johnston, 1997; Baddeley, 2000, 2006). Das Modell von Baddeley besteht aus vier Hauptkomponenten: der zentralen Exekutive, der phonologischen Schleife, dem visuellen Notizblock und dem episodischen Puffer. Die zentrale Exekutive ist dabei als Überwachungssystem tätig, welches den Informationsfluss steuert und für die Kontrolle und Regulierung kognitiver Prozesse verantwortlich ist. Die phonologische Schleife sowie der visuelle Notizblock und episodischen Puffer sind für die kurzfristige Speicherung von Informationen verantwortlich (Baddeley, 2000, 2006). Kinder mit mathematischen Schwierigkeiten haben häufig ein Defizit in einem Aspekt der zentralen Exekutive (Holmes & Adams, 2006). Passolunghi und Siegel (2001) untersuchten den Unterschied zwischen guten und schlechten Problemlösern. Sie unterscheiden sich durch die unzureichende Unterscheidungsfähigkeit zwischen relevanten und weniger relevanten Informationen. Deutlich schlechtere Problemlöser haben zudem Schwierigkeiten, irrelevante Informationen zu hemmen, was ihre Arbeitsspeicherkapazitäten überanstrengt. Eine ähnliche Auffassung vertreten Andersson und Lyksell (2007), die die gleichzeitige Koordination von Verarbeitung und Speicherung für die schlechteren Leistungen von Schülern verantwortlich machen. Durch

die gleichzeitige Koordination wird die Ausführung einer arithmetischen Aufgabe gestört, was zu einer längeren Lösungszeit und zu mehr Fehlern führt. Die Koordination ist bei mathematischen Problemstellungen häufig nötig, wenn Zwischenergebnisse gespeichert werden müssen und verschiedene Verfahrensschritte durchgeführt werden sollen. Kinder mit Schwierigkeiten der Arbeitsspeicherkapazität haben auch im Klassenzimmer Schwierigkeiten (Gathercole & Pickering, 2000), gleichzeitig Informationen zu hören, zu verarbeiten, zu speichern und die neu erworbenen Informationen in ein bestehendes Wissenssystem zu übertragen. Auch bestätigen die Autoren, dass diese Kinder Schwierigkeiten haben, die numerischen und visuellen Informationen gleichzeitig zu verarbeiten und zu speichern.

2.1.2.2 Mathematische Fähigkeiten in kognitiven Strukturmodellen

Im Folgenden soll ein Blick auf die mathematischen Fähigkeiten aus der Sicht der Strukturforschung kognitiver Fähigkeiten geworfen werden. Die Strukturforschung kognitiver Fähigkeiten ist eher an den Ergebnissen der oben beschriebenen Prozessschritte interessiert und bezeichnet mathematische Fähigkeiten als erfolgreich oder eben nicht erfolgreich gelöste Aufgaben aus der Mathematik (Li & Schmiedeck in Brunner, 2006). Deshalb sollen nun Strukturmodelle vorgestellt werden, die insbesondere mathematische Fähigkeiten berücksichtigen. In dem grundlegenden Modell von Spearman (1904, 1927), das bereits im Rahmen der Beschreibung kognitiver Modelle erläutert wurde, resultieren mathematische Fähigkeiten auf einem unter dem g-Faktor der allgemeinen kognitiven Fähigkeiten angeordneten „problem arithmetic“-Faktor, während einfache Additionsaufgaben den spezifischen Faktoren zugeordnet werden konnten. Thurstones Alternativmodell (1938) zu Spearman sah mehrere gleichrangige Primärfaktoren der Intelligenz vor. In Bezug auf mathematische Fähigkeiten sind hier insbesondere die Rechenfertigkeit (numerical), und das quantitative Reasoning zu nennen. Letzteres beruht auf einer, von Gustafsson und Undheim (1992) als „arithmetical reasoning“-Faktor bezeichneten Fähigkeit, die Thurstone selber nicht explizit interpretierte (Thurstone, 1938). Aus der nachfolgenden Erweiterung von Primärfaktoren wurde die Entwicklung hierarchischer Faktormodelle vorangetrieben, da die Primärfaktoren nicht exakt voneinander unterscheidbar waren und teils andere Fähigkeiten mit erklärten. Faktoren einer ersten Ordnung erklären in diesen hierarchischen Modellen die Faktoren zweiter Ordnung und ggf. einer dritten Ordnung. In der Entwicklung des Gf-Gc-Modells kristallisierter und fluider kognitiven Fähigkeiten auf Basis des ursprünglich von Cattell entwickelten Modells wurde in der Weiterentwicklung von Horn und Noll die generelle mathematische Fähigkeit (gq) „quantitative knowledge“ formuliert. Nach Horn und Noll

werden unter letzterer das Verständnis und die Anwendung mathematischer Konzepte verstanden, die zum einen durch eine generelle Mathematikfähigkeit und durch mathematikspezifische Fähigkeiten erklärt wird.

In der nachfolgenden Konzeption von Carroll (1993) ist dagegen eine generelle kognitive Fähigkeit den fluiden Denkfähigkeiten (*gf*), den kristallisierten Fähigkeiten (*gc*) und der Informationsgeschwindigkeit (*gs*) übergeordnet. Das quantitative Reasoning ist dabei den fluiden Fähigkeiten zugeordnet, während Carroll die Rechenfertigkeiten der Informationsgeschwindigkeit zuordnet, da er diese als einfache Abarbeitung von Rechenregeln versteht, bei denen eher die Geschwindigkeit – „cognitive speed“ – entscheidend sei. Dieses Modell genießt weitreichende Akzeptanz und wurde umfassend empirisch validiert. Für die stoffgebietsspezifische Unterscheidung sprechen die theoretischen Überlegungen und empirisch validierten Befunde der konzeptionellen Überlegungen von Jäger et al. (1997) aus dem Berliner Intelligenzstrukturmodell sowie dem *gf-gc*-Modell von Horn-Noll (1997) und Cattell (1963). Auch in dem Drei-Stratum-Modell von Carroll (1993) wird eine hierarchische Unterscheidung mathematischer Fähigkeiten nach Stoffgebieten vorgenommen. Für quantitatives Reasoning werden nach Carroll (1993) Fähigkeiten aus Algebra, Arithmetik und Geometrie benötigt.

Das Verständnis, welchen Bereichen mathematische Fähigkeiten den kognitiven Modellen zugeordnet werden, ist durchaus unterschiedlich. So wird aus den angeführten Modellen ersichtlich, dass Vernon eine generelle numerische Fähigkeit konzeptionalisiert, die den stoffgebietsspezifischen mathematischen Fähigkeiten übergeordnet ist. In der *Gf-Gc*-Theorie von Cattell (1963) und Noll und Horn (1997) werden mathematische Fähigkeiten als Problemlösefähigkeit zur Bewältigung mathematischer Anforderungen verstanden, während im Berliner Intelligenzstrukturmodell eine generelle mathematische Fähigkeit extrahiert wird. In dem Berliner Intelligenzstrukturmodell von Jäger et al. (1997) wurde die mathematische Fähigkeit aus der Verarbeitungskapazität und der numerischen Fähigkeit theoretisch konstruiert. Die mathematische Leistung resultiert dabei aus der operativen Fähigkeit, der generellen mathematischen Fähigkeit und der allgemeinen kognitiven Fähigkeit.

Carroll wiederum sieht keine generelle mathematische Fähigkeit, sondern ordnet die mathematischen Fähigkeiten dem quantitativen Reasoning, worunter schlussfolgerndes Denken innerhalb mathematischer Beziehungen verstanden wird, unter dem Anspruch der Problemlösungsfähigkeit zu. Weiter unterscheidet er die Rechenfertigkeit, bestehend aus

den mathematischen Stoffgebieten Arithmetik, Algebra und Geometrie, womit er den Umgang mit Zahlen (Zählen von Objekten, Erkennen von Zahlen) und arithmetische Operationen (Addition, Subtraktion, Multiplikation, Division) mit Zahlen, Brüchen oder Dezimalzahlen meint. Hier sind einfache Prozeduren anzuwenden, bei denen es auf Korrektheit und Geschwindigkeit ankommt. Deshalb ordnet er sie der Informationsgeschwindigkeit zu.

Brunner bemerkt im Vergleich der angeführten Modelle, dass die Platzierung der Stoffgebiete zu den entsprechenden Fähigkeiten recht unterschiedlich erfolgte. Daher sind für eine eindeutigere Zuordnung neuere mehrdimensionale Faktorenmodelle geeignet, generelle kognitive Fähigkeiten darzustellen und zugleich auch bereichsspezifische kognitive Faktoren aufzunehmen Brunner (2006). Die sowohl generelle mathematische Fähigkeit als auch stoffgebietsspezifische mathematische Fähigkeiten lassen sich mittels faktorenanalytischer Strukturmodelle ebenso präzise abbilden. Im Vergleich der Modelle mathematischer Fähigkeiten wird deutlich, dass die Gemeinsamkeit darin besteht, dass sie hierarchisch aufgebaut sind und spezifische mathematische Teilfähigkeiten einer globalen mathematischen Fähigkeit unterordnen. Die oben beschriebenen Modelle können allerdings mathematische Fähigkeiten nicht in ihrer Komplexität erfassen. Sie sind nicht in der Lage, die spezifischen Beziehungen der mathematischen Leistungen und der kognitiven Fähigkeiten angemessen zu repräsentieren.

Bei dem Versuch, beide Konstrukte unabhängig voneinander zu messen, wird das Problem offensichtlich und löst die Diskussion um die Identität der Konstrukte mathematischer Schulleistungen und kognitiver Fähigkeiten aus. In diesem Zusammenhang ist vor allem die Aufgabenähnlichkeit präsent. So wies Rindermann (2006) bereits darauf hin, dass Intelligenztestaufgaben kaum von mathematischen Aufgaben zu unterscheiden sind. Andererseits werden in der pädagogisch-psychologischen Literatur die Konstrukte als getrennte Konstrukte betrachtet (Renkl & Stern; 1994, Spada & Wichmann, 1996), während wiederum in der PISA-Rahmenkonzeption zur Messung mathematischer Problemlöseprozesse und der Formulierung der Ansprüche, was in kognitiven Testverfahren erfasst werden soll, deutlich wird, dass *beide* Anforderungen Problemlösungen erfordern. Damit wird die Unterscheidung von Items, die quantitatives Reasoning (nach Carroll, 1993) erfassen und Aufgaben, wie sie in kognitiven Fähigkeitstests erfragt werden, schwierig. So erfordert das quantitative Reasoning, laut Definition die problemorientierte Anwendung von Algebra, Geometrie und Arithmetik. Auch die Unterscheidung nach Neuartigkeit des Wissens erscheint nach Brunner (2006) nicht plausibel, da auch dann nicht klar werde, ob es sich um

konzeptionelles Wissen (im mathematischen Sinne) oder um fluide Fähigkeiten (im kognitiven Sinne) handelt. Auch die von Cooley und Lohnes (1976) vorgeschlagene Unterscheidung nach funktionalem Einsatz wird kritisiert. Die funktionale Unterscheidung sieht eine Bezeichnung nach ihrer Fragestellung vor. Wenn ein Test eingesetzt wird, um die aktuelle Kompetenz zu messen, handelt es sich um einen Fähigkeitstest, wenn aber die Frage der Prognose gestellt wird, handelt es sich um einen Befähigungstest. Bond (1989) zitiert diese Ansicht: „Yesterday’s achievement is today’s ability and tomorrow’s aptitude.“ (Bond, 1989 in Brunner, 2006). Snow (1989) beispielsweise lehnt die Unterscheidbarkeit der Konstrukte gänzlich ab, da Schulleistungstests sowohl Wissensstrukturen erfragen als auch Strukturierungs- und Organisationsprozesse erfordern, die wiederum den Wissenserwerb erleichtern. Wenn also Schulleistungstests die Problemlösefähigkeit erfassen, so sind diese Fähigkeiten ebenso in Intelligenztests erforderlich, da das Organisieren, Generalisieren, Adaptieren und die Anwendung von Wissen genau dies erfordern. Die oben erläuterten Schwierigkeiten thematisiert Brunner (2006), der feststellt, dass eine historische Trennung ebenso wenig möglich ist wie die psychometrische Trennung des dem Testverfahren zugrundeliegenden Messmodells. Er konzipierte daraufhin ein Modell, das die spezifischen hierarchischen Beziehungen der mathematischen Leistungen und der kognitiven Fähigkeiten angemessen repräsentiert. Das Modell wurde aus den Modellannahmen von Vernon (1964), dem Berliner Intelligenzstrukturmodell von Jäger et al. (1997) und Carrolls Drei-Stratum-Modell (1993) abgeleitet. Aus diesen Strukturtheorien kognitiver Fähigkeiten suchte Brunner (2006) zu erklären, welche kognitiven Fähigkeiten mathematische Leistungsunterschiede erklären können. Das von Brunner validierte Nested-Faktor-Modell zur Struktur mathematischer Schülerleistung zeigt, dass interindividuelle Unterschiede durch Unterschiede in der allgemeinen kognitiven Fähigkeit, einer mathematikspezifischen Fähigkeit sowie einer stoffgebietsspezifischen Fähigkeit, die den schulischen Wissenserwerbsprozessen zugeordnet wird, erklärt werden können. Diese drei erwähnten Fähigkeiten interagieren. Die empirische Prüfung der Theorie kristallisierter und fluider Fähigkeiten (Horn & Noll, 1997) hielt der empirischen Prüfung stand (Brunner, 2006), aber es zeigte sich eine schlechtere Modellpassung als das von Brunner vorgeschlagene Nested-Faktor-Modell, welches im Gegensatz zu dem Modell kristallisierter und fluider Denkfähigkeiten auch eine allgemeine kognitive Fähigkeit berücksichtigt. Eine stoffgebietsunabhängige Fähigkeit zum generellen Lösen mathematischer Probleme wird als mathematikspezifische Fähigkeit verstanden, wie sie auch bereits von Geary (1994) interpretiert wird. Sie setzt sich

aus der Fähigkeit zur Repräsentation mathematischer Probleme (Meyer, 1985) und der Arbeitsgedächtniskapazität zusammen. Erstere besteht aus der Fähigkeit, problemrelevante Informationen aus der Aufgabenstellung oder gegebenen Abbildungen zu extrahieren und mittels mathematischen Wissens in ein mentales mathematisches Problemmodell zu integrieren. Erforderlich sind hier schematisches, strategisches und prozedurales Wissen, das stoffgebietsspezifisch erworben wurde. Die Prozesse zum Aufbau der mentalen mathematischen Modelle sind jedoch nicht stoffgebietsspezifisch, sondern resultieren aus der generellen mathematikspezifischen Fähigkeit und sind auch unabhängig von den allgemeinen kognitiven Fähigkeiten (Brunner, 2006). Diese wiederum erklären Unterschiede der interindividuellen Leistung durch Unterschiede in der Arbeitsspeicherkapazität (Geary, 1993, 2007). Die stoffgebietsspezifischen Fähigkeiten repräsentieren also das stoffgebietsspezifische Wissen, die mathematikspezifische Fähigkeit gilt als Fähigkeit zur Repräsentation stoffgebietsunabhängiger Fähigkeit zum mathematischen Problemlösen, während die allgemeine kognitive Fähigkeit die Arbeitsspeicherkapazität beinhaltet (Brunner, 2006). In der mathematikspezifischen Fähigkeit sind allerdings nach Brunner wiederum auch Prozesse des Textverstehens wirksam, was zu einer gewissen unscharfen Abgrenzung von kognitiven Fähigkeiten führt. Brunner spricht sich für die präzise Modellierung der beteiligten Konstrukte aus, da nur bei entsprechend genauer Operationalisierung eine genaue Darstellung des Zusammenwirkens verschiedener Faktoren möglich ist (2006). Und so resümiert Brunner, dass eine globale mathematische Fähigkeit von einer allgemeinen kognitiven Fähigkeit und einer stoffgebietsspezifischen Fähigkeit zu unterscheiden sei (2006). Neben der Frage der Unterscheidung zwischen kognitiven Fähigkeiten und mathematischen Fähigkeiten steht die Frage, ob und wie sich kognitive Fähigkeiten und mathematische Schulleistungen vorhersagen lassen, im Zentrum des Forschungsinteresses.

2.1.2.3 Prädiktoren von Mathematikleistungen

Im Interesse der Forschungsbemühungen zu den Zusammenhängen zwischen kognitiven Fähigkeiten und mathematischen Schulleistungen steht vor allem die Beantwortung der Frage, wie sich beide Konstrukte vorhersagen lassen. Genaues Wissen über die Zusammenhänge zwischen mathematischem Kompetenzerwerb und den kognitiven Fähigkeiten sind für die Forschung und die pädagogische Praxis wichtig. Auch die Beantwortung der Frage, woher die Leistungen beider Bereiche resultieren, also die Frage nach der Ursache, in der rückblickenden Betrachtung, ist zentral, insbesondere hinsichtlich der frühzeitigen Intervention bzw. Prävention von Leistungsproblemen. Die Vorhersage von mathematischen

Leistungen ist nicht minder schwierig (Byrnes & Wasik, 2009), da sie von der Motivation oder dem Lehrerverhalten beeinflusst werden und auch Vorkenntnisse eine große Rolle spielen. Der empirische Nachweis ist aber oft methodisch schwierig, da nur eine Teilmenge möglicher Einflussfaktoren in die Analysen einbezogen werden kann. Dies führt zur Überschätzung der einflussnehmenden analysierten Faktoren (Byrnes & Wasik, 2009). So ergeben sich teilweise ganz unterschiedliche Ergebnisse. In diesem Kapitel sollen die für diese Arbeit wesentlichen Forschungsergebnisse vorgestellt werden, um den Forschungsstand darzulegen bzw. um auf noch offene Fragen hinzuweisen. Es existiert eine Reihe von Längsschnittstudien zu den multifaktoriellen Einflussfaktoren, in denen versucht wurde, die Leistungsergebnisse für Mathematik vorherzusagen.

Flanagan und McGrew (1997) führten auf Grundlage der *gf-gc*-Theorie Studien zu der prädiktiven Gültigkeit von Schulleistungsmaßen aufgrund von spezifischen kognitiven Fähigkeiten durch und fanden heraus, dass kognitive Faktoren nach der CHC-Theorie geeignet waren, den deutlichen Zusammenhang zu den mathematischen Schulleistungen festzustellen. Sie stellten fest, dass kristallisierte Intelligenz moderate Beziehungen mit Arithmetik und mäßige bis starke Beziehungen mit mathematischem Schlussfolgern aufwies. Fluide Intelligenz (*Gf*), das Kurzzeitgedächtnis (*Gsm*) und der Arbeitsspeicher hatten ihrerseits moderate Beziehungen mit den Mathematikleistungen. In einer ähnlichen Studie von McGrew, Flanagan, Keith und Vanderwood (1997) konnte anhand der Spezifikation mittels Strukturgleichungsmodellierung gezeigt werden, dass *Gf* und *Gc* starke Prädiktoren für mathematisches Schlussfolgern sind. Zusätzlich zu den Untersuchungen der Beziehungen zwischen den allgemeinen kognitiven Fähigkeiten und der akademischen Leistung sollte die Rolle der verbalen Fähigkeiten bei der Vorhersage der Schulleistungen ermittelt werden. Schulische Leistungen sind nach der CHC-Theorie als kristallisierte Intelligenz kategorisiert. Die Forschung zeigte, dass verbale Fähigkeiten, durch kognitive Tests gemessen, als sehr gute Prädiktoren für schulbezogene Leistungen herangezogen werden können (Flanagan, Andrews & Genshaft, 1997). In einer recht umfassenden Modellierung von (Byrnes & Wasik, 2009) wurden zahlreiche Faktoren analysiert, wie die Leistungsmotivation, das Alter, Unterrichtsfaktoren, das Unterrichtsniveau und psychologische Umgebungsfaktoren wie die Klassenstruktur und die Gleichaltrigenumgebung. Die Autoren weisen darauf hin, dass die Häufigkeit, mit den Lerninhalten konfrontiert zu werden, eine entscheidende Rolle für spätere Mathematikleistungen spielt. Insofern plädieren sie dafür, dass bereits im Vorschulalter die Voraussetzungen zum Erwerb bestimmter Niveaustufen ange-

glichen werden müssten, um Benachteiligungen von Kindern mit schlechten Voraussetzungen aufzuheben. Damit interpretieren sie einen bedeutenden prädiktiven Einfluss des Vorwissens auf spätere Mathematikleistungen im Vergleich zu anderen (nicht kognitiven) Faktoren. Zu den einflussnehmenden kognitiven Faktoren auf die Mathematikleistungen zählt auch das Arbeitsgedächtnis (Cragg & Gilmore, 2014), das eine Wirkung auf den Faktenerwerb besitzt. Das Arbeitsgedächtnis wiederum wirkt auch auf den Konzepterwerb mathematischer Fähigkeiten sowie auf die Steuerung von Prozeduren beim Lösen von mathematischen Problemen (Cragg & Gilmore, 2014). Als unspezifischer Prädiktor mathematischer Leistungen wirkt die Gedächtniskapazität, da ein schneller Zugriff auf Zahlworte im Langzeitgedächtnis mehr als 25 % der Varianz im Bereich der numerischen Basisfertigkeiten bestimmt (Krajewski & Schneider, 2006; Swanson, 2008). Als Gedächtniskapazität werden die Zeit, für die Informationen im Arbeitsgedächtnis aktiviert bleiben können und die Geschwindigkeit der Informationsverarbeitung bezeichnet. Das Arbeitsgedächtnismodell nach Baddeley (1986, 2006) bildet die theoretische Grundlage. Es besteht aus drei Komponenten: der phonologischen Schleife, dem visuell-räumlichen Notizblock und der zentralen Exekutive. Der phonologischen Schleife und dem visuell-räumlichen Notizblock ist die zentrale Exekutive übergeordnet. In ihr werden Informationen aus der phonologischen Schleife und dem visuell-räumlichen Notizblock koordiniert und überwacht (Kontrolle der Aufmerksamkeit, Abruf von Wissen aus dem Langzeitgedächtnis, Unterdrückung von Handlungsimpulsen, Koordination von Informationen). So werden beispielsweise bei Multiplikationsaufgaben der Zehnerübergang (Thomas et al., 2006) von der zentralen Exekutive gesteuert. Routinierter Umgang mit mathematischen Aufgaben durch Übung und automatisierte Abläufe im Rechnen entlasten die zentrale Exekutive (Ricken & Fritz, 2006) und können interindividuelle Unterschiede in den mathematischen Leistungen erklären (Grube & Barth, 2004). Im Kontext der mathematischen Leistungen werden in der phonologischen Schleife auditive bzw. sprachgebundene Informationen, wie Aufgabeninformationen oder Zwischenergebnisse, gespeichert, verarbeitet und abgerufen. Ihre Bedeutung liegt insbesondere beim Aufbau mathematischen Wissens (Grube, 2006). In dem visuell-räumlichen Notizblock werden statistische Merkmale (Formen, Bewegungssequenzen aus visuellen und räumlichen Informationen) gespeichert. Die Beziehungen zwischen den Arbeitsspeicherkomponenten und Mathematikleistungen werden auch von Fris van den Bos et al. (2013) und Geary et al. (2004) betont, wobei darauf hingewiesen wird, dass die Arbeitsspeicherkapazität domänenspezifisch die mathematische Leistungsfähigkeit beeinflusst. Allerdings

wurden diese Ergebnisse auf korrelativer Ebene ermittelt. Kinder mit frühen soliden Vorerfahrungen erzielen bessere mathematische Leistungen (Krajewski & Schneider, 2009). Grundlegende mathematische Fähigkeiten werden von frühem mathematischem Verständnis bereits im Kindergartenalter beeinflusst (Krajewski & Schneider, 2006). Es existieren auch Hinweise, dass die mathematischen Vorkenntnisse, die im Vorschulalter angeeignet werden, von Krajewski und Schneider (2006) mathematische Vorläuferfertigkeiten genannt, mathematische Leistungen besser vorhersagen als die Intelligenz (Weißhaupt, Peucker & Witz, 2006), wenn das Arbeitsgedächtnis einbezogen wird (Krajewski et al., 2008). Als Vorläuferfertigkeiten gelten das Zahlvorwissen, das Zählwissen sowie erste Rechenfertigkeiten (Krajewski, 2008).

Dennoch wird der Einfluss der Intelligenz nicht gänzlich abgelehnt, sondern sie besitzt einen indirekten Einfluss, vor allem auf das numerische Vorwissen (Krajewski & Schneider, 2006; Krajewski et al., 2008). Die Intelligenz zeigt ab der zweiten Klassenstufe allerdings einen abnehmenden Einfluss, während die Bedeutung des Vorwissens zunimmt und fehlendes Vorwissen kaum mehr durch intellektuelle Fähigkeiten kompensiert werden kann (Weinert, 1997). Unsicherheiten im Vorläuferwissen führen demnach zu Schwierigkeiten bei der Festigung neu erlernten mathematischen Wissens und beim Abruf aus dem Langzeitgedächtnis treten häufig Fehler auf (Gaupp et al., 2004). Die Leistungsvarianz über die Grundschulzeit bleibt relativ stabil (Stern, 1998; Weinert & Helmke, 1997). Die Bedeutung von Intelligenz und Vorwissen hinsichtlich der Mathematikleistungen wurde bereits von Renkl und Stern (1994) hervorgehoben. Helmke und Weinert (1997) wiederum betonen, dass Kinder mit höheren kognitiven Fähigkeiten effektivere Problemlösestrategien entwickeln und dadurch Vorteile hinsichtlich des weiteren Wissensvorsprungs erwerben. Jedoch schrumpft der Vorteil der Intelligenz im Laufe der Schulzeit durch die zunehmende Bedeutung des Wissens und des Vorwissens (Helmke & Weinert, 1997). Beispielsweise können Multiplikationsaufgaben erst dann erfolgreich gelöst werden, wenn Subtraktions- und Additionsfähigkeiten zuvor sicher erworben wurden (Krajewski, 2003). Allerdings wird von Geary, Brown und Samaranayke (1991) eingewandt, dass es Kindern bei Beeinträchtigungen der Gedächtniskapazität schwerfalle, gleichzeitig benötigte Informationen im Arbeitsgedächtnis zu behalten. Dies führe zu längeren Lösungszeiten und einer höheren Fehleranzahl. Auch Krajewski (2004) betont die Bedeutung von Arbeitsgedächtnisleistungen im Zusammenhang mit den Rechenleistungen.

2.1.2.4 Messung und Bewertung von schulischen Leistungen

Nachdem nun in den vorhergehenden Abschnitten die Modelle und Theorien mathematischer Schulleistungen vorgestellt wurden, soll nunmehr das Augenmerk auf Messung und Bewertung schulischer Leistungen gelegt werden. Wie von Jensen (1998) festgestellt, gibt es individuelle Lernunterschiede. Manche Menschen können unter den gleichen Bedingungen des Lernens schneller lernen als andere. Darüber hinaus sind einige nicht in der Lage, die gleichen Dinge, die von anderen gelernt werden, zu lernen. Normalerweise werden Kinder, die mehr als ihre Altersgenossen aus den gleichen Erfahrungen lernen, als überdurchschnittliche oder hohe Leistungsträger betrachtet. Ebenso werden Kinder, die weniger als ihre Altersgenossen aus denselben Erfahrungen lernen, als leistungsschwach betrachtet.

Bildungssysteme haben in der Regel ein Bewertungsverfahren, das an Standards orientiert ist, um das Lern- bzw. Leistungsniveau der Schüler zu testen. Solche Curriculum-Standards sind Teil der meisten Bildungssysteme und Schüler sollen Wissen und Erfahrungen erwerben, um diese Standards zu erreichen. Eine Bewertung der Qualität der Arbeit eines Schülers, um seine oder ihre Bildungsleistung zu beurteilen, kann entweder eine Gesamtleistung oder eine fachspezifische Leistung, wie Lesen oder Mathematikleistungen, sein. Die Messung der individuellen Unterschiede und des Leistungsniveaus entscheiden unter anderem über die Initiierung pädagogischer Interventionen oder die weitere Schullaufbahn.

Bildungsleistungen werden mit einer Vielzahl von Methoden beurteilt. Schulabschlüsse sowie Abschlussnoten und Lehrerauswertungen können zur Bewertung der Leistung herangezogen werden. Da Schulleistungen stärker lehrplangebunden sind, fließen Anstrengung, Gewissenhaftigkeit sowie Lehrerurteilskompetenz als irrelevante Variablen mit in die Notengebung ein. Standardisierte Leistungstests dagegen messen spezifische Bildungsleistungen (Hesse & Latzko, 2008). Deshalb werden Schulleistungstests als ein Maß für die Schulleistung angesehen (Hesse & Latzko, 2008). Obwohl standardisierte Schulleistungstests nicht garantieren, dass alle Schüler vollkommen objektiv und losgelöst von ihrem Lernkontext beurteilt werden können, minimieren Schulleistungstests potenzielle Störvariablen wie Lehrerstil und Lehrerwahrnehmungen (Kaufman, 2012; auch Hesse & Latzko, 2008). Der Vorteil der standardisierten Testverfahren ist, dass sie auf Faktorenanalysen beruhen und testübergreifend angewendet werden können, sodass die Beurteilung der gemeinsamen Varianz über die Tests hinweg möglich ist und damit die Fehlervarianz minimiert werden kann (Kaufman, 2012). Die Korrelation zwischen dem Intelligenzquotienten

und der akademischen Leistung ist typischerweise höher beim Einsatz standardisierter Leistungstests als wenn lediglich Schulnoten verglichen werden (Mackintosh, 2011).

Die Evaluation der angestrebten Realisierung der im Lernort Schule zu realisierenden Bildungsstandards erfolgt über die Leistungserfassung. Dabei soll zunächst eine Vergleichbarkeit der Leistung von Schülerinnen und Schülern in den einzelnen Schulen sichergestellt werden. Die Leistungsbewertung wird aber auch als Instrument der Lernberatung und Lernförderung verstanden und soll auch als Form der Rückmeldung zu den Lernfortschritten auf Schülerebene beitragen und deren Anstrengungsbereitschaft unterstützen. Die Kultusministerkonferenz empfiehlt in diesem Zusammenhang ein Kompendium zum Thema Schulleistungsmessung und verweist auf die Publikation von Weinert (2001), in der dieser die Grundlagen schulischer Leistungsmessung analysiert. Darin benennt Weinert eine ausreichende schulische Lerneffektivität und deren regelmäßige interne und gegebenenfalls externe Evaluation als notwendige Form der Leistungserfassung, um die Vergleichbarkeit der Lernfortschritte zu sichern (Weinert, 2001).

Was tatsächlich aber als Leistungsmessung in der Schule definiert wird, ist häufig unklar. In der Diskussion darum, ob fachliche oder überfachliche Leistung im Sinne von Wissen oder kognitiven Kompetenzen bzw. Handlungskompetenzen erfasst werden sollen, gehen die Meinungen auseinander (Weinert, 2001). Während ein Diskussionsstrang die geringe Bedeutung fachlicher Leistung betont, erachten Vertreter der gegenteiligen Position jedoch die fachlichen Kompetenzen als wesentlich, da sie doch die Voraussetzung schaffen, Probleme im sozioökonomischen Kontext des Lebensalltags lösen zu können. Dafür sei das notwendige fachliche Wissen eine Voraussetzung, um Transferaufgaben überhaupt erst lösen zu können. Der Mehrzahl der Schulklassen stehen kaum zuverlässige Vergleichsdaten zur Beurteilung des eigenen Leistungsstandes zur Verfügung (Weinert, 2001). Um dieser Tatsache Rechnung zu tragen, wurden fachspezifische Tests entwickelt, die den Gütestandards schulischer Leistungsmessung genügen (Ingenkamp, 2008; Lukesch, 1998; Hesse & Latzko, 2011).

2.1.2.5 Schulleistungstests

Schulleistungstests sind ein wissenschaftliches Routineverfahren zur Feststellung des Kenntnisstands in einem oder mehreren inhaltlich spezifizierten kognitiven Lehrzielbereichen, dabei werden Aussagen über die Leistungshöhe aufgrund des Vergleichs mit den Leistungen einer zumeist gleichaltrigen Vergleichsgruppe bzw. Klassenstufe formuliert (Lukesch, 1998). Schulleistungstests werden im Rahmen pädagogischer Diagnostik eingesetzt,

um Voraussetzungen und Bedingungen von Lehr- und Lernprozessen zu ermitteln und zu analysieren (Ingenkamp, 1996). Im Vordergrund steht das Erkenntnisbemühen, wobei allerdings die Gründe für den Einsatz variieren können. So wird die Wirkung von schulstrukturellen Maßnahmen evaluiert, oder auch die Wirkung von Fördermaßnahmen evaluiert. Schulleistungstestverfahren sind standardisiert und im Gegensatz zu beiläufigen Urteilen beruhen sie auf objektiven Ergebnissen. Mit den gewonnenen diagnostischen Informationen können auf individueller Ebene das Leistungsniveau von Schülern ermittelt und damit geprüft werden, ob der Schwierigkeitsgrad von Aufgaben den Leistungsmöglichkeiten eines Schülers entspricht. Auf der interindividuellen Ebene können Unterschiede im Leistungsniveau innerhalb einer Klasse aufgedeckt werden, um den Unterricht zu optimieren. Ferner können auf institutioneller Ebene objektive Zeugnisse erstellt werden, oder auch die Leistungen einer Klasse mit den Leistungen einer anderen Klasse verglichen werden, auch über Bundesländer hinweg. Im Rahmen der vorliegenden Arbeit werden die genannten Schulleistungstests eingesetzt, da sie objektive Ergebnisse zum individuellen Leistungsniveau in verschiedenen mathematischen Bereichen ermitteln und diese dann einem interindividuellen Vergleich dienen.

Zur Messung mathematischer Schulleistungen sind die Testverfahren der DEMAT-Reihe von Krajewski, Küspert und Schneider (2002), Krajewski, Liehm und Schneider (2004) sowie Roick, Gölitze und Hasselhorn (2004, 2006) geeignet, da sie die erreichten Leistungen klassenstufenspezifisch erfassen können. Dabei orientieren sie sich an dem klassenstufenspezifischen Curriculum sowie an den von der ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland vorgeschlagenen Kompetenzstufen (Krajewski et al., 2004). Dazu zählen das Problemlösen, die Analyse mathematischer Probleme, die Generierung von Lösungsstrategien und die Übertragungsfähigkeit auf weitere mathematische Probleme und auch das Kommunizieren von mathematischen Lösungswegen. Dazu zählt auch das Argumentieren, worunter die Erkenntnis und die Prüffähigkeit des Wahrheitsgehaltes mathematischer Aussagen verstanden wird, weiterhin das Modellieren, das die Fähigkeit zur Abbildung und Lösung von Sachaufgaben beinhaltet, aber auch das Darstellen, welches für die Lösung mathematischer Probleme erforderlich ist.

2.1.3 Zwischenresümee kognitive Fähigkeiten und mathematische Schulleistungen

In der Zusammenfassung der bisherigen Befunde bleibt festzuhalten, dass für die Beantwortung der Eingangsfrage eine akribische Spurensuche notwendig ist. Zum einen macht die

Vielzahl der Definitionen von Intelligenz sowie die Vielzahl der dazugehörenden Theorien die zweifelsfreie Eingrenzung der Begriffe nicht leicht. Im Rahmen dieser Arbeit wurde aber mit der Erweiterung des Terminus Intelligenz auf den Terminus kognitive Fähigkeiten eine Definition des für diese Arbeit wichtigen Konstrukts gefunden, der der komplexen Thematik Rechnung trägt. So werden kognitive Fähigkeiten im Rahmen dieser Arbeit als dynamisch und multidimensional sowie als Potential und Ressource verstanden (Perleth, 1997). Intelligenz im Sinne kognitiver Fähigkeiten wird also im Zusammenhang der vorliegenden Studie als eine umfassende, komplexe Fähigkeit zur Lösung von Problemen definiert, die sich dadurch auszeichnet, dass sie sich ändernde Anforderungen bewältigen kann. Kurz gesagt, Intelligenz wird hier verstanden als das Maß, wie wir mit den Herausforderungen unserer Umwelt zu Recht kommen (Wittmann, Eisenkolb & Perleth, 1997).

Unter der Vielzahl der Theorien zur Intelligenz konnte das CHC Modell mit besonderer Betonung auf der Investitionstheorie von Cattell als das für diese Arbeit geeignetste vorgestellt werden. Die zugrunde liegende Gf-Gc-Theorie postuliert fluide und kristallisierte Intelligenzfaktoren. Obwohl fluide Denkfähigkeiten und kristallisierte Denkfähigkeiten nach der Investitionstheorie von Cattell (1971) als getrennte Konstrukte verstanden werden, so werden sie doch als zwei Aspekte eines gemeinsamen g-Faktors betrachtet. Nach Cattell ist die fluide Intelligenz eine Hauptursache für die Leistung, da sie die Kapazität bedeutet, effizienter und schneller zu lernen. Demnach spielt die fluide Denkfähigkeit eine relevante Rolle für das Lernen und hilft den Schülern, neue Fähigkeiten und Fertigkeiten zu erwerben, die erforderlich sind, um sich neue schulische Wissens Elemente anzueignen (Cattell, 1971; Primi et al., 2010). Insofern wird die fluide Intelligenz als ein Faktor des Lernens betrachtet, der sich vor allem in neuartigen Situationen bemerkbar macht und die Aneignung von Fähigkeiten und Kenntnissen erleichtert. Zudem erleichtert die kristallisierte Intelligenz auch die Erfassung komplexer Probleme. Dieses Potential lässt sich leichter in Erfahrung und Wissen umwandeln und kristallisiert quasi die Intelligenz. So wird die Vermittlungsrolle von bisherigen schulischen Leistungen und Intelligenz gestützt. In ihrer Erweiterung um Umweltmoderatoren wie soziale Einflüsse wie Familie und Schule sowie Persönlichkeitsfaktoren wie Motivation und Ausdauer in dem Modell von Perleth und Sierwald kann die Investitionstheorie von Cattell dazu dienen, die für diese Studie relevanten Fragen theoretisch zu untermauern. Eine besondere Bedeutung kristallisierte sich in diesem Zusammenhang für den Faktor Vorwissen heraus, denn es konnte eruiert werden, dass Wissen in dieser Form sowohl die kognitiven Fähigkeiten als auch die schulischen Leistungen beeinflusst. Dabei sind die bereichsspezifischen Kenntnisse von enormer Bedeutung und zunehmend

weniger von allgemeinen intellektuellen Fähigkeiten determiniert (Weinert & Helmke, 1987). Nicht unerwähnt bleiben darf in diesem Zusammenhang die beachtliche Rolle, die die Zugehörigkeit zu einer bestimmten Klasse und das damit verbundene Unterrichtsangebot spielt. Kognitive Fähigkeiten spielen aber dennoch eine bedeutende Rolle, wenn Problemlösefähigkeiten (Ackerman, 1988, 1989; Weinert & Helmke, 1988) erforderlich sind oder neues Wissen erworben wird. Insofern konnten die bisher vorliegenden Ergebnisse zu den Zusammenhängen zwischen Mathematikleistungen und kognitiven Fähigkeiten bestätigt werden. Allerdings – auch das soll nicht unerwähnt bleiben – enthält die kognitive Entwicklung zumindest in der hier interessierenden mittleren Kindheit individuelle Veränderungsspielräume, die die Validität langfristiger Vorhersagen prinzipiell beschränken dürften. Diese generelle Schlussfolgerung ist aber eher eine plausible theoretische Überlegung als eine empirisch bereits hinreichend belegte Aussage.

Eine Sichtung der zur Messung von Intelligenz existierenden Testverfahren machte anschließend deutlich, dass ihre Anwendung im Rahmen dieser Arbeit nur unter dem Vorbehalt der messtheoretischen Überprüfung sinnvoll erscheint, um Messinvarianzen und anderen messspezifischen Fehlern, die Aussagekraft einiger der gesichteten Studien mindern, vorzubeugen.

Die Definition von mathematischen Schulleistungen, dem zweiten, für diese Studie wichtigen Konstrukt, und seine Verortung in existierenden Modellen der Strukturforschung machte deutlich, dass kognitive Fähigkeiten und schulische Leistungen autonome Variablen sind, die allerdings beide gemeinsam prädiktiv für die Schulleistungen sind. Diese Annahme geht konform mit mehreren Studien, die die Identifizierung von Intelligenz als relevanten Prädiktor für akademische Ergebnisse bestätigen (z. B. Soares, 2015; Deary et al., 2007; Karbach et al., 2013; Lemos et al., 2013; Lynn & Vanhanen, 2012; Primi et al., 2010; Weber et al., 2013).

Auf Basis dieser aus dem aktuellen Forschungsstand gewonnenen Erkenntnisse ist die nun folgende Formulierung der vier Forschungsfragen zwingend, um die bisher weder theoretisch noch empirisch zweifelsfrei geklärten Beziehungen zwischen den beiden Konstrukten aufzuzeigen.

2.2 Ableitung der Fragestellungen

Werden kognitive Fähigkeiten und mathematische Fähigkeiten differenziert betrachtet, ist bislang ungeklärt, inwiefern beide Konstrukte womöglich identische Fähigkeiten erfassen.

Über die faktorielle Differenziertheit der Intelligenz und schulischer Leistungen wurden erhebliche Debatten geführt (siehe Flanagan & McGrew, 1997). Diese Debatten gehen von drei Möglichkeiten aus: Vorstellungen, dass es sich bei Mathematikleistung und Intelligenz um überlappende Konstrukte handelt, werden bei Ceci (1991), innerhalb Deutschlands auch von Rindermann (2006), in der Gegenargumentation wiederum von Baumert, Brunner, Lüdtke und Trautwein (2007) heftig diskutiert. Eine Übersicht geben Watkins et al. (2007). Von einer starken Beeinflussung von Mathematikleistungen durch Intelligenz gehen hingegen Lubinski und Dawis (1992) aus. Aber auch Annahmen einer gegenseitigen Beeinflussung von Mathematikleistungen und kognitiven Fähigkeiten im Kindes- und Jugendalter sind ebenso formuliert worden (Stanovich, 1986; Schneider & Stefanek, 2004). Die vorliegende Arbeit versucht daher mit der ersten Fragestellung einen grundlegenden theoretischen Beitrag zur Klärung bisheriger kontroverser theoretischer Annahmen zu leisten: Erfolgt die Erfassung kognitiver Fähigkeiten und der Mathematikleistungen innerhalb der Messzeitpunkte partiell unabhängig? Inwiefern handelt es sich folglich bei Mathematikleistungen und kognitiven Fähigkeiten um unabhängige Konstrukte?

Wenn – wie sich herauskristallisiert hat – kognitive Fähigkeiten und schulische Leistungen autonome Variablen sind, die jedoch beide gemeinsam letztlich die Schulleistungen vorherzusagen vermögen, stellt sich nun die Frage nach den bisher ungeklärten Zusammenhängen zwischen beiden. Im Rahmen der Arbeit wird daher die komplexe Verwebung von kognitiven Fähigkeiten und mathematischen Schulleistungen im Grundschulalter betrachtet. Daher klärt die zweite Fragestellung zunächst, inwieweit die Messung kognitiver Fähigkeiten und der Mathematikleistungen über die Zeit vergleichbar ist.

Das Interesse gilt jedoch auch der Stabilität des Niveaus kognitiver Leistungen, welches ab dem frühen Erwachsenenalter eine hohe Niveaustabilität erreicht (Ziegler, 2004; Rost, 2010; Schneider & Hasselhorn, 2007). Für das Grundschulalter scheint die Struktur kognitiver Leistungen ab dem 10. Lebensjahr annähernd stabil zu sein (Rost, 2010; Trost, 1998; Franić, 2014), wobei der schulischen und familiären Lernumwelt sowie den nicht kognitiven Persönlichkeitsmerkmalen (z. B. Perleth, 1997) eine große Bedeutung hinsichtlich der weiteren Steigerung, insbesondere hinsichtlich der Positionsveränderungen im Vergleich mit Gleichaltrigen beigemessen wird. Im Zentrum der dritten Fragestellung dieser Arbeit soll daher beantwortet werden, inwieweit die erfassten Konstrukte kognitive Fähigkeiten und mathematische Leistungen im Grundschulalter als stabil zu betrachten sind.

Etwas außerhalb der bisherigen Forschungsbemühungen stand die Frage nach der gegenseitigen Beeinflussung von Schulleistungsmerkmalen und kognitiven Fähigkeiten.

Der Verlauf kreuzverzögerter Effekte von kognitiven Fähigkeiten auf spätere Mathematikleistungen unter gleichzeitiger Berücksichtigung der gegenteiligen Effekte (von früheren Mathematikleistungen auf kognitive Fähigkeiten) ist noch nicht hinreichend untersucht worden und so werden theoretische Annahmen über die zeitliche Präzedenz zwischen den genannten Konstrukten immer noch kontrovers diskutiert. Bislang berichtete Längsschnittuntersuchungen umfassen nur zwei Messzeitpunkte und lassen nur ungenaue Aussagen zu dem zeitlichen Verlauf kreuzverzögerter Effekte zu. Die vierte Fragestellung wird daher wie folgt formuliert: Sind vor allem kognitive Fähigkeiten prädiktiv für Mathematikleistungen oder kann von einer gegenseitigen Beeinflussung ausgegangen werden?

2.2.1 Fragestellung 1 – Faktorielle Differenziertheit

Aufgrund der offensichtlich hohen Zusammenhänge zwischen Intelligenz und schulischen Leistungsmaßen plädierte Spearman (1904) für die Annahme, dass kognitive Fähigkeiten und schulische Bildungsmaße identisch seien. Spearmans Ergebnissen folgte eine theoretische Auseinandersetzung über den Zusammenhang zwischen Intelligenz- und Bildungsleistungen, die zu einer der zentralen Fragestellung in der Intelligenzforschung wurde. So widmeten sich auch Binet und Simon (1916) der Thematik, inwiefern Bildungsabschlüsse anhand kognitiver Fähigkeiten vorhergesagt werden können. Der Ausgangspunkt ihrer Untersuchungen waren die von Spearman festgestellten hohen Korrelationen, aufgrund derer angenommen wurde, dass Bildungsleistungen nichts Anderes als kognitive Fähigkeiten messen. Anlass für die Folgeuntersuchungen durch Binet und Simon (1916) war der letztlich nicht vollständig geklärte Nachweis, dass vornehmlich der nach Spearman benannte allgemeine Intelligenzfaktor g (Spearman, 1904, 1927) die Bildungsleistungen bestimmt.

In dieser Debatte, wie kognitive Fähigkeiten und (mathematische) schulische Leistungen zusammenhängen, konnte bis heute keine Einigkeit erzielt werden. Drei verschiedene Positionen kristallisierten sich heraus, die Basis einer heftig geführten Theoriekontroverse sind. Einige Forscher vertreten die Position, dass aufgrund der teilweise hohen Ähnlichkeit der Aufgaben zur Messung der Konstrukte davon ausgegangen werden kann, dass eine Unterscheidung ohnehin nicht möglich ist und die Konstrukte also als identisch zu betrachten sind. Andere gehen von einer Überlappung der Konstrukte kognitive Fähigkeiten und mathematische Schulleistungen aus, während Vertreter der dritten Position die Meinung vertreten, dass beide Konstrukte klar voneinander unterscheidbar und daher differenziert zu betrachteten seien.

Das folgende Kapitel soll daher einen Einblick in den aktuellen theoretischen Diskussionsstand dieser drei Richtungen geben. Zunächst soll auf die Positionen eingegangen werden, die die Identität der Konstrukte annehmen, um danach auf die Vertreter der Überlappungstheorie einzugehen. Abschließend wird die Position, die die Unabhängigkeit der Konstrukte annimmt, dargestellt.

2.2.1.1 Forschungsposition – Identität

Ausgehend von Spearmans Ergebnissen und den Bemühungen von Binet und Simon (1916), die Prognose von Bildungsabschlüssen aufgrund der kognitiven Voraussetzungen zu ermöglichen, wurden die Aufgaben in Intelligenztests und Mathematikleistungstests hinsichtlich ihrer Ähnlichkeit untersucht. Tatsächlich ist in der Debatte um die Konstruktidentität von schulischen Leistungen und dem Intelligenzkonstrukt zunächst die augenscheinliche Aufgabenähnlichkeit prominent. In der Tat werden auch heute in aktuellen kognitiven Fähigkeitstests und in Schulleistungstests häufig gleiche Elemente oder Aufgaben verwendet. So setzten etwa Ditton und Kreckler (1995) die quantitative Subskala des KFT 4-13 von Heller, Gaedicke, Weinländer (1976) als Messinstrument von Mathematikleistungen ein. Rindermann weist zudem darauf hin, dass in zahlreichen Intelligenztests arithmetische Subtests ein fester Bestandteil seien (Rindermann, 2006). Die Aufgabenähnlichkeit hat zur Folge, dass oftmals für die Lösung von Aufgaben in kognitiven Fähigkeitstests gleiche Fähigkeiten erforderlich sind, wie sie für die Lösung von mathematischen Aufgaben benötigt werden. So sind häufig zur Bearbeitung von Mathematikschulleistungstests auch Lese- und Textverständnis erforderlich. Beispielsweise müssen im deutschen Mathematiktest für die 4. Klassenstufe von Roick, Gölitze und Hasselhorn (2006) die Fragestellungen inhaltlich richtig verstanden werden, um die Aufgabe im Bereich Geometrie erfolgreich zu bewältigen. Das erfordert Textverständnis.

Auch von anderen Forschern wie Gottfredson (2003), Herrnstein und Murray (1994) oder auch Rost (2005) wird argumentiert, dass inhaltliche Aufgaben sehr ähnlich sind und zur Lösung von Mathematik- und auch von Intelligenzaufgaben Problemlösefähigkeiten, Allgemeinwissen und Schulwissen erforderlich seien, die sich kaum empirisch abgrenzen lassen. Auch, so die Forscher, liegen den Leistungen in Intelligenztests und schulischen Tests vergleichbare Kompetenzen, wie schlussfolgerndes Denken, Abstraktionsfähigkeit und Verständnis zugrunde. So habe die kognitive Verarbeitungsgeschwindigkeit direkten Einfluss auf die Intelligenz und beeinflusse folglich die Schulleistung, wie Rindermann anführt (2006). Die Verarbeitungsgeschwindigkeit wird dabei als eine grundlegende geistige

Fähigkeit verstanden, die eine wichtige Ressource der Intelligenz darstellt. (Rindermann & Neubauer, 2004). Auch Ceci (1991) macht geltend, dass die Inhalte von schulischen Leistungstests und die Inhalte kognitiver Eignungstests, wie sie zurzeit aufgebaut sind, sehr ähnlich und sowohl theoretisch als auch statistisch untrennbar verwandt seien. Tatsächlich ist der erfolgreiche Abruf früherer Informationen zum Einsatz aktueller Anforderungen wiederum von Prozessen der Informationsverarbeitung (Jensen, 1998) und von der Fluidität der Denkfähigkeit im Sinne Cattells abhängig. Die genannten Kompetenzen sind somit als kognitive Voraussetzungen für die erfolgreiche Bewältigung beider Aufgabenkonzeptionen erforderlich.

Anhaltenden Grund für die Identitätsposition liefern auch die deutlichen korrelativen Zusammenhänge zwischen der generellen kognitiven Fähigkeit und Schulnoten, die mit Korrelationen von .40 bis .70 (Mackintosh, 1998) angegeben werden. Ähnlich hohe Korrelationen in Höhe von .50 berichtete auch Jensen (1998). Auch in den Untersuchungen von (Naglieri & Bornstein, 2003) weisen Intelligenz- und Leistungstestergebnisse Korrelationskoeffizienten in den Bereichen von .60 bis .70 auf. Naglieri und Bornstein untersuchten in einer Metaanalyse die Zusammenhänge zwischen Intelligenz und Leistung. Sie sichteten verschiedene Testmanuale und Zeitschriftenartikel in Bezug auf Korrelationen zwischen Intelligenz und Leistungstests zum einen und zwischen Intelligenz und Leistungssubtests zum anderen. Sie kamen zu dem Ergebnis, dass Intelligenz mit Leistungssubtests um .65 korreliert. Es wurden 13 Studien mit mehr als 200 Testpersonen (zwischen 1284 und 1940) und 6 Studien mit weniger als 200 Personen (zwischen 28 und 198 Teilnehmer) auf ihren Zusammenhang zwischen Intelligenz und schulischen Leistungsmaßen analysiert. Die minimale Korrelation war .33 und die maximale .89. Die mittlere Höhe betrug .61. Aber die Gruppe der kleinen Stichproben hatten Korrelationen von .36 bis .64, während die der großen Studien von .63 bis .74 reichte. Dies, so die Autoren Naglieri und Bornstein (2003), zeigt eindrucksvoll den zweifelhaften Wert kleiner Stichproben. Die Ergebnisse einschränkend halten die Autoren fest, dass die Daten teils zu gleichen, teils zu unterschiedlichen Zeitpunkten erhoben wurden. Darüber hinaus überlappen sich einige Testverfahren im Konstruktverständnis und nicht alle Daten waren normalverteilt. Zudem wiesen viele Studien methodische Schwächen auf. So fanden die Autoren über verschiedene Testverfahren hinweg häufig dieselben Fragen in einem Lesetest sowie in einem verbalen Intelligenztest. In beiden wurde nach einem anderen Wort für „big“ bzw. „large“ gefragt. Daher verwundert eine hohe Korrelation nicht. Die Autoren plädieren für die Testkonzeption, die auf der Trennbarkeit der Konstrukte beruht. Die Überzeugung, dass schulische Leistungsmaße und

kognitive Fähigkeiten ähnliche Fähigkeiten erfassen, daher vollkommen korrelieren und somit identische Konstrukte seien, teilen auch Frey und Detterman (2004) sowie Koenig, Frey und Detterman (2008). Anhand einer Stichprobe von 1075 Teilnehmern, die einen Zulassungstest für das amerikanische College durchführten, konnten Koenig, Frey und Detterman (2008) Korrelationen mit dem allgemeinen Intelligenzfaktor in Höhe von .77 feststellen. Auch zeigten sich mit einer kleineren Stichprobe von 149 Personen Korrelationen von .61, die die Zusammenhänge zwischen Raven-Matrizen-Tests und Teilleistungen aus dem Bildungsleistungstest darlegen. Die Autoren schlussfolgern, dass der allgemeine Intelligenzquotient anhand der Ergebnisse des Bildungsleistungstests vorhergesagt werden könne. Linda Gottfredson zieht für den Beleg einer Äquivalenz von Intelligenz und schulnahen kognitiven Fähigkeiten den US-amerikanischen National Adult Literacy Survey (NALS) heran, der ähnlich wie PISA das Leseverständnis erhebt. Ihrer Meinung nach werde vor allem *g* oder „verbales *g*“ gemessen, da der Umgang mit komplexen verbalen Informationen, schlussfolgerndem Denken und der Informationsverarbeitung erhoben werde. Die verschiedenen Subdimensionen seien untereinander hoch korreliert und die prognostische Validität ähnlich hoch wie bei Intelligenztests. Zudem werde diese Argumentation auch dadurch gestützt, dass die prognostische Validität von Intelligenztests und Tests, die schulisch-kognitive Leistungen erheben, länderübergreifend ähnlich hoch mit dem späteren Einkommen und der beruflichen Position korreliert seien (Gottfredson, 2003).

Auch Koenig, Frey und Detterman (2008) kamen zu dem gleichen Ergebnis, als sie den Zusammenhang zwischen Leistungstests und Intelligenztests anhand einer nationalen amerikanischen Datenbasis mit 1075 Teilnehmern aus dem Jahr 1979 untersuchten. Sie erfassten eine allgemeine kognitive Fähigkeit *g*. Die Ergebnisse zeigen eine Korrelation des Collegezulassungstests mit dem allgemeinen Fähigkeitstest in Höhe von .77. In einer neueren Stichprobe mit 149 Teilnehmern zeigten sich Korrelationen des Raven-APM-Tests mit dem Bildungsleistungstest (ACT) in Höhe von .61. Die Autoren folgern, dass der amerikanische Colletest eingesetzt werden könne, um die kognitiven Fähigkeiten der allgemeinen Bevölkerung vorherzusagen.

Im deutschsprachigen Raum dominieren diese Debatte hauptsächlich Rindermann (2006) und Rindermann und Neubauer (2004), die sich für eine identische Betrachtung kognitiver Fähigkeiten und schulischer Mathematikleistungen aussprechen. Sie analysierten internationale Schulleistungsstudien hinsichtlich ihrer inhaltlichen Aufgaben. Ihr Resümee unterstützt die Annahme, dass Intelligenztestverfahren und Schulleistungstestverfahren ähnliche Konstrukte erheben. Rindermann begründet seine Schlussfolgerung mit der

Feststellung, dass Leseleistungsaufgaben auch Wahrnehmungsfähigkeiten erfordern. Ferner seien häufig in Leseleistungsaufgaben Tabellen- und Grafikinterpretationsfähigkeiten erforderlich. Auch Aufgaben aus den Bereichen der Naturwissenschaften verlangen umfangreiches Textverständnis, so Rindermann. Für das Lösen von Mathematikaufgaben, die sich an Alltagssituationen orientieren, sei ebenfalls häufig Textverständniskompetenz sowie allgemeines Denken notwendig. Vornehmlich allgemeines Denken ähnelt den Problemlöseaufgaben, welche wiederum häufig in Intelligenztests eingesetzt werden. Aufgrund der hohen Korrelationen zwischen den intelligenznahen Skalen in Mathematikleistungstests und den inhaltlichen Aufgabenbereichen in Intelligenztests wird die Verschiedenartigkeit der Tests von einigen Intelligenzforschern angezweifelt.

Ein Blick auf die in dieser Arbeit verwendeten Aufgaben der DEMAT- und KFT-Autoren bestätigt diese Sichtweise und wird an dem folgenden Beispiel deutlich.

„Hier sollst du zeigen, wie gut du etwas wiedererkennen kannst. Im ersten Bild siehst du ein Schneemännchen oder ein Strichmännchen von vorne. In einem von den Bildern rechts neben dem schwarzen Balken siehst du das Männchen von hinten. Nur ein Bild entspricht dem ersten Männchen. Welches ist es? Kreuze dieses bitte an!“

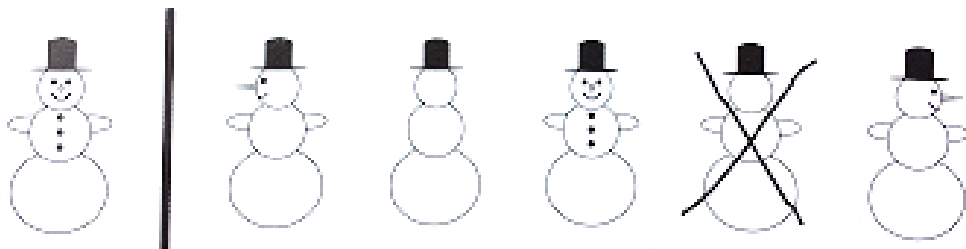


Abbildung 6 Beispielaufgabe DEMAT aus Roick, Gölitze & Hasselhorn (2006).

Werden wiederum Aufgaben des Verbalbereichs des KFT 1-2 (Abou-Koura & Perleth, 2005) betrachtet, so sind z. B. in der Aquariumaufgabe auch Zählfähigkeiten zur Beantwortung notwendig:

„Seht euch die Bilder in der ersten Reihe an. Es gibt vier Bilder. Unter jedem Bild ist ein kleiner Antwortkreis. Bei jeder Frage sollt ihr einen Antwortkreis ausfüllen. Damit gebt ihr an, dass das eure Antwort sein soll. Jetzt kommt die

Frage 1: Welches Bild zeigt den Schmetterling?

Frage 2: Welches Bild in der nächsten Reihe zeigt ein Aquarium, in dem ein Fisch ist?

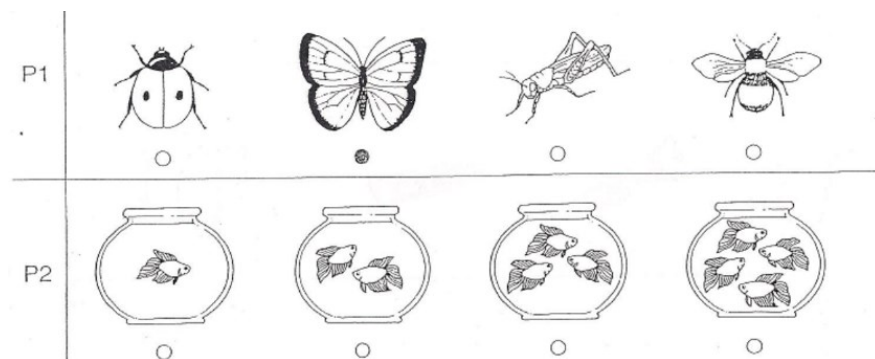


Abbildung 7 Beispielaufgabe KFT aus Abou-Koura & Perleth (2005).

Auch in Mathematikleistungstests sind wiederum die Fähigkeiten zur Interpretation von Grafiken und schlussfolgerndem Denken im Sinne der Plausibilitätsprüfung erforderlich, die eine Abgrenzung auf rein mathematische/quantitative Bereiche erschweren. Die folgende Grafik zeigt ein Beispiel aus dem Mathematikleistungstest für die erste Klassenstufe im Bereich Arithmetik.

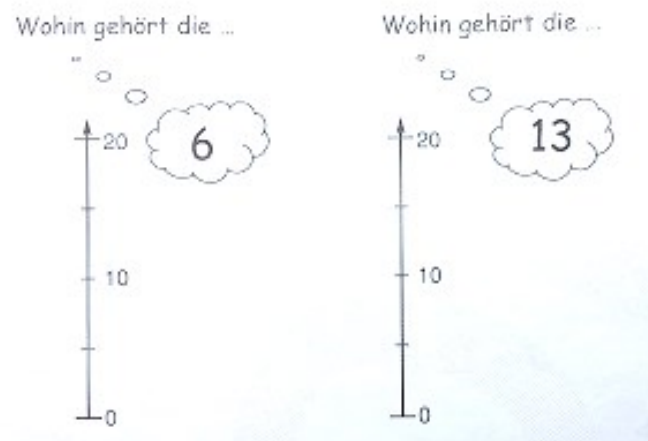


Abbildung 8 Beispielaufgabe DEMAT 1 aus Krajewski, Küspert & Schneider (2002).

Auch in Sachrechenaufgaben sind explizit Problemlösefähigkeiten gefordert.

„In mein Rechenheft male ich ein Rechteck. Die eine Seite ist 7 Kästchen lang. Die andere Seite ist 4 Kästchen lang. In jedes Kästchen des Rechtecks male ich eine Blume.“

Frage: *Wie viele Blumen male ich in das Rechteck?*
Antwort: *Es sind _____ Blumen.*“

Abbildung 9 Beispielaufgabe DEMAT 1 aus Krajewski, Küspert, & Schneider (2002).

Aus der Ähnlichkeit der Aufgaben, die auch oben deutlich wird, resultiert, so Rindermann (2007), eine unklare empirische Beziehung der Skalenkorrelationen untereinander, die in Schulleistungstests häufig unzureichend berichtet wird (Rindermann, 2007). Rindermann (2006) kam aufgrund seiner Untersuchungen zu der Schlussfolgerung, dass eine erfolgreiche Bearbeitung von Aufgaben, die Problemlösungsstrategien erfordern, nicht ausschließlich durch erworbenes Wissen gelingen kann. Diese Aufgaben erfordern ebenso induktives und deduktiv-logisches, schlussfolgerndes Denken sowie Abstraktionsfähigkeit, tieferes Verständnis und auch die Fähigkeit zu der Herstellung von Beziehungen und struktureller Zusammenhänge. Rindermann (2004) argumentiert, dass diese Fähigkeiten unter dem Begriff allgemeine Denkfähigkeit subsumiert werden können, womit die Argumentation für einen prominenten Faktor *g* der allgemeinen kognitiven Intelligenz spricht.

Die Identität der Konstrukte werde auch über die Ähnlichkeit der Definitionen deutlich, so Rindermann (2006). Erforderliche Problemlösefähigkeiten, wie sie in Mathematikschulleistungstests erfasst werden, ähneln selbst in der Definition dem Konstrukt der Intelligenz (OECD, 2004b). So ähneln beispielsweise die Definitionen der mathematischen Grundbildung den Definitionsdimensionen, mit denen intellektuelle Fähigkeiten erfasst werden. Die Definition der mathematischen Grundbildung wird dabei relativ unkonkret beschrieben. Dies resultiert offenbar aus der Schwierigkeit, einzelne Anforderungsbereiche voneinander abzugrenzen. Auch die zahlreichen und teils sehr unterschiedlichen Definitionen von kognitiven Fähigkeiten lassen eine gewisse Unkonkretheit aufgrund der schwierigen Abgrenzung einzelner Bereiche vermuten (Rindermann, 2006). Auch andere Autoren teilen diese Überzeugung. So sagen zum Beispiel auch Naglieri und Bornstein (2003), dass einige Studien einer klareren Definition von Intelligenz bedürfen. Kaufmann et al. (2012) weisen ebenfalls darauf hin, dass das Verständnis allgemeiner kognitiver Fähigkeiten und Bildungsleistungsmaßen teilweise unscharf definiert ist. Manche Definitionen, die Grundlage für Intelligenztestverfahren sind, fokussieren nur auf den allgemeinen Intelligenzfaktor, die fluide Intelligenz (Carroll, 1993; Gustafsson, 1988). Dieser Fakt wird von Baumert et al. (2007) als Gegenargument zur Identitätsposition herangezogen. Sie gehen davon aus,

dass eine eingrenzende Definition von kognitiven Fähigkeiten auf schlussfolgerndes und abstraktes Denken zu kurz greift (Baumert et al., 2007). Darunter werde hauptsächlich der Umgang mit Problemstellungen verstanden, für die kein spezifisches Wissen zur erfolgreichen Bearbeitung verfügbar ist. Sie stützen sich auf Carroll (1993) und Gustafsson (1984) und sind der Meinung, damit werde hauptsächlich fluide Intelligenz beschrieben, die beispielsweise mittels sprachfreier Matrizen- oder figuraler Aufgabenbereiche von kognitiven Fähigkeitstests erfasst werden könne. Solche Aufgaben finden sich in internationalen Schulvergleichsstudien, sind aber tatsächlich von vielfältigen weiteren Faktoren, wie dem Anregungsgehalt der Umwelt und der Qualität der Lernumgebungen abhängig (Ceci, 1991; Helmke & Weinert, 1997) und werden darüber hinaus von nicht-kognitiven Persönlichkeitsmerkmalen (Perleth, 1997) moderiert.

In der zuvor schon näher beleuchteten Studie von Deary et al. (2007) wurde die allgemeine kognitive Fähigkeit im Sinne von *g* über 25 akademische Fächer hinweg verglichen. Aufgrund der dort gemessenen extrem hohen Korrelationen zwischen dem latenten allgemeinen Intelligenzfaktor und den latenten Merkmalen von Bildungsleistungen schlagen Deary et al. vor, dass die Konstrukte als nahezu identisch eingeschätzt werden. Da nur etwa die Hälfte der Varianz der Bildungsleistungen durch den allgemeinen Intelligenzfaktor statistisch erklärt werden kann, verwiesen Deary et al. allerdings auf die notwendige Berücksichtigung von Messfehlern oder auch anderen Einflussfaktoren, wie Schulbesuchsdauer, Motivation oder Engagement, Persönlichkeitsfaktoren und Anstrengungsbereitschaft sowie elterliche Unterstützung oder Unterrichtsqualität. Von einer Beeinflussung von Mathematikleistungen durch Intelligenz gehen deshalb Lubinski und Dawis (1992) aufgrund der Schwierigkeit der Abgrenzung spezifischer kognitiver Fähigkeiten aus. Das veranschaulichten sie auch in ihrem Modell *Organisation of Human Abilities* (Lubinski & Dawis, 1992).

Es lässt sich also zusammenfassend feststellen, dass vorrangig die bekannt hohen Korrelationen einigen Forschern dazu Anlass geben, die Identität der Konstrukte anzunehmen und somit den Einfluss anderer Faktoren als eher gering einzuschätzen.

2.2.1.2 Forschungsposition – Faktorielle Differenziertheit

Trotz der nachgewiesenen hohen Korrelationen lassen sich aber auch Argumente finden, die nicht für eine Identität der Konstrukte sprechen. So wenden Kaufman et al. (2012) ebenso wie Ceci (1991) ein, dass die Identität der Konstrukte nicht abschließend geklärt sei und dass hohe Korrelationen noch keine Kausalinterpretation zulassen. Und obwohl auch

Jensen (1998) feststellte, dass schulische Leistung und kognitive Fähigkeiten stark korrelieren, ist er dagegen, aus dem einen Konstrukt direkt auf das andere zu schließen. Die Leistungen des jeweils anderen Konstrukts seien nicht direkt vorherzusagen, da es zum Beispiel Umweltmoderatoren oder personeninterne nicht-kognitive und kognitive Faktoren zu berücksichtigen gebe. Kaufman et al. (2012) kritisieren deshalb, dass einige Forscher Intelligenz und akademische Leistung gleichsetzen. Er betont die Bedeutung der schulischen Wissensvermittlung auf Prozesse kognitiver Leistungen und hebt den daraus resultierenden, nicht zu überschätzenden Einfluss von Schule als bedeutendem Ort der Förderung kognitiver Fähigkeiten hervor. Ceci geht sogar so weit, intellektuelle Fähigkeiten im Sinne des allgemeinen Intelligenzverständnisses nach Spearman (1904) als ein Ergebnis schulisch erworbener Leistung (Ceci, 1991) und Schulleistungen und kognitive Fähigkeiten folglich als überlappende Konstrukte zu bezeichnen. Durch die langfristige Orientierung kann schulische Bildung nach Ceci (1991) daher selbst als Förderungselement zukünftiger schulischer Leistungen und damit als Förderungssituation für kognitive Fähigkeiten fungieren (Ceci, 1991). Von einer hohen Überlappung geht Ceci auch deshalb aus, weil zahlreiche kognitive Fähigkeitstests eben auch schulisch erworbene Fähigkeiten abfragen. Ceci nennt beispielsweise den kognitiven Fähigkeitstest von Heller und Perleth (2000). So weisen die in der TIMSS erhobenen mathematischen Leistungen überwiegende Ähnlichkeit mit der ersten quantitativen Subskala des KFT auf (TIMSS). Die Konzeption des kognitiven Fähigkeitstests beruht jedoch auf dem CogAT (Cognitive-Ability-Test) von Thorndike und Hagen (1978) und nimmt voneinander trennbare, spezifische Konstrukte an (Perleth, 1997, 2007; Heller & Perleth, 2000).

Ackerman, Beier und Boyle (2005) hingegen widmeten sich den personeninternen nicht-kognitiven Faktoren. Sie untersuchten, ob das Arbeitsgedächtnis, wenn es integraler Bestandteil der kognitiven Fähigkeiten ist, mit der allgemeinen Intelligenz identisch oder nahezu identisch ist. In einer Metaanalyse von 86 zugrundeliegenden Studien konnten sie belegen, dass das Arbeitsgedächtnis nicht identisch mit allgemeinen kognitiven Fähigkeiten ist. Ackerman, Bayer und Boye (2005) sprechen sich aufgrund ihrer Forschungsbefunde deutlich gegen die Existenz eines allgemeinen, alleinigen g-Faktors aus.

Im Zusammenhang mit den personeninternen kognitiven Faktoren soll nun ein näherer Blick auf diese Forschungsperspektive geworfen werden. Eine Übersicht zu dieser Dauerdebatte liefern hier Watkins et al. (2007). Sie belegten in ihren Studien den prädiktiven Einfluss von kognitiven Variablen auf Schulleistungsmaße, während der umgekehrte

Einfluss nicht zu belegen war. Sie interpretieren somit, dass kognitive Fähigkeiten und Leistungsmaße keine identischen Konstrukte sind (Kaufman, 2012).

Die Ergebnisse von Deary et al. (2007), die eine Korrelation zwischen den Konstrukten in Höhe von .81 nachwiesen, lassen letztlich keine eindeutige Antwort auf die Frage nach der Unabhängigkeit vs. Identität der Konstrukte zu. Deary et al. selbst betonen den potentiellen Einfluss anderer nicht-kognitiver Faktoren, die gegen eine vollständige Identität schulischer Leistungsmaße und allgemeiner Intelligenzfaktoren sprechen. Darüber hinaus geben sie Messfehlereinflüsse als Quelle einer fehlenden eindeutigen empirischen Beantwortung an.

Kaufman et al. (2012) bestätigen in ihren Untersuchungen der Beziehungen zwischen kognitiven Fähigkeiten und Schulleistungsmaßen ebenfalls, dass keine Identität zwischen beiden vorliegt, wenngleich sie ähnlich hohe Korrelationskoeffizienten wie Deary et al. (2007) fanden. Sie konnten dieses Ergebnis über zwei unterschiedliche Testbatterien konsistent belegen. Darüber hinaus konnten sie Effekte fluider Intelligenz auf kristallisierte Intelligenz im Sinne sich gegenseitig verstärkender positiver Effekte nachweisen. Kaufman et al. (2012) beschließen ihre Untersuchungen der Beziehungen zwischen kognitiven Fähigkeiten und Schulleistungsmaßen mit der Aussage, dass keine Identität zwischen beiden vorliegt, sondern von einer Überlappung oder Unabhängigkeit ausgegangen werden muss.

Letztlich bleibt festzuhalten, dass eine starke Korrelation zwischen Leistungsfaktoren und Intelligenzfaktoren existiert, jedoch längst noch keine vollständige Korrelation. Ebenso wenig, so Kaufman (2012), seien aber auch Dearys (2007) Ergebnisse perfekt korreliert, obwohl sich bei Kaufman auf latenter Ebene zwischen allgemeinen kognitiven Faktoren (COG-g) und Schulleistungsmaßen (ACH-g-Faktoren) starke Korrelationen zeigen. Im Gegensatz zu Deary, der aus seinen Resultaten Identitätsbelege ableitet, nennt Kaufman (2012) die weniger gut messbaren, nicht-kognitiven Variablen der Leistung als Ursache für den Mangel an perfekter Korrelation (2012). Sie sind theoretisch und empirisch anhand zahlreicher Schulleistungsbedingungsmodelle (beispielsweise Heller, 1991; Perleth, 1997; Ceci, 1991) belegt. Solche einflussnehmenden Faktoren können motivational sein oder auch von anderen Variablen, wie Trainingszeit oder Instruktionsqualität (Carroll, 1963; Haertel, Walberg & Weinstein, 1983; Heller, 1991; Perleth, 1997) moderiert werden. Kaufman et al. konzipieren ACH-g wegen der Bereichsspezifität als zusammengesetzte Variable und plädieren ebenso wie Soares (2015) aufgrund ihrer Ergebnisse für die Beibehaltung des getrennten Verständnisses beider Konstrukte, indem sie auf das von McGrew et al. konzipierte

und validierte CHC-Modell kognitiver Fähigkeiten verweisen. Im Zuge der Untersuchungen von Soares zur Vorhersage von schulischen Leistungen von Schülern der 9. Klassenstufe durch die kognitiven Fähigkeiten und die schulische Leistung zu Beginn der Untersuchung in Klassenstufe 7, konnte er mittels Testverfahren zur Erfassung numerischer, verbaler Erfassungsdimensionen die Unabhängigkeit der Prädiktoren nachweisen. Die indirekte Wirkung der anfänglichen akademischen schulischen Leistungen erklärt einen Großteil der endgültigen schulischen Leistungen. Damit zeigt Soares den Einfluss beider Konstrukte als unabhängige Prädiktoren. Doch auch die direkte Wirkung der Intelligenz auf die anfängliche schulische Leistung war signifikant und erklärte etwa 50 % der anfänglichen Schulnoten zu Beginn der Klassenstufe 7, während die direkte Wirkung von Intelligenz auf die Schulleistungen in der 9. Klassenstufe nicht signifikant war. Die Ergebnisse von Soares lassen die Annahme zu, dass sowohl kognitive Fähigkeiten als auch schulische Leistungen autonome Variablen sind, die jedoch beide gemeinsam letztlich die Schulleistungen vorherzusagen vermögen.

McGrews Ergebnisse sprechen ebenfalls für die Annahme getrennter, spezifischer Fähigkeiten. Zum einen gäbe es trotz hoher Korrelationen methodische Probleme bei der Erfassung der spezifischen Erfassung. Daneben verbleibe ein recht hoher Anteil eben nicht aufgeklärter Varianz. Die Autoren führen auch pädagogische Gründe an, die bei einer Interpretation als identisches Konstrukt in Beratungssituationen deutliche Nachteile bergen können, da Schulleistungen kumulativ bedingt seien und von Motivationen, Attributionen, nicht-kognitiven Persönlichkeitsmerkmalen und weiteren Faktoren beeinflusst werden. Dafür sprechen auch die Befunde von Ceci (1991) oder Perleth (1997).

Abschließend wird die Position der Autoren in den Fokus gestellt, die dem geänderten Verständnis des g-Faktors Rechnung tragen. Während Spearman g noch als eine ganz allgemeine Fähigkeit zum Denken identifizierte, hat sich inzwischen das bereichsspezifische Verständnis von kognitiven Fähigkeiten mehr und mehr durchgesetzt (Flanagan, Keith, McGrew & Vanderwood, 1996; McGrew, Vanderwood, Flanagan & Keith, 1996). Richtungsweisend ist hier die von Horn und Noll erweiterte CHC-Theorie. Sie spielt auch in der Diskussion um die Identität der Konstrukte eine große Rolle. Die Untersuchungen von McGrew, Flanagan, Keith und Vanderwood (1997) und Alfonso, Flanagan und Radwan (2005) sind hier federführend. Die Autoren begründen die Notwendigkeit, spezifische Fähigkeiten zu erfassen, weil diese zum Beispiel für die diagnostische Erfassung spezifischer Mathematikfähigkeiten nötig sind. Trotz der offensichtlich ähnlichen Beziehung beider

Konstrukte (Korrelationen von bis zu .83) ist also von einem getrennten Verständnis auszugehen. Von der Position, dass die spezifische Fähigkeitserfassung wesentlich präzisere Aussagen erlaubt, ging auch Perleth (1997) aus, der quantitativ-numerische Fähigkeiten annimmt (auch Perleth, 1997; Heller & Perleth, 2000) und diese demnach nicht unter der Vorstellung allgemeiner kognitiver Fähigkeiten subsumiert. Die Konzeption des in der vorliegenden Arbeit eingesetzten kognitiven Fähigkeitstests in den Versionen KFT 1-2R, KFT 3R, KFT 4-12R von Abou-Koura und Perleth (2005), Perleth und Heller (2008) und Heller und Perleth (2000) beruht auf dieser Annahme. In diesem Zusammenhang widmeten sich Johnson et al. (2007) in einer Replikationsstudie der Frage, inwieweit Leistungstests mit Intelligenztestverfahren übereinstimmen und ob sich aufgrund der hohen Korrelation schließen lässt, dass beide Konstrukte identischer Natur sind. Sie verwendeten fünf Testverfahren, die vornehmlich Wahrnehmungsfähigkeiten im Fokus hatten, sowie eine kleine Auswahl aus dem Cattell Culture Fair Test (1978) und setzten sie an 500 erwachsenen Personen, die in den Niederlanden lebten, ein. Sie wollten wissen, inwieweit sich bestätigen lässt, dass hauptsächlich ein allgemeiner Intelligenzfaktor g für die Qualität der Aufgabenbewältigung entscheidend ist. In Übereinstimmung mit der früheren Studie (2004) konnten sie belegen, dass die meisten Intelligenztestverfahren zwar einen Basiswert im Sinne von g erfassen, der sich konsistent und exakt über mehrere Testverfahren hinweg erfassen lasse. Zugleich aber werden auch spezifische Faktoren gemessen, die von dem Konzept der höheren Ordnung im Sinne von g nicht gemessen werden. Ihre Erfassung bleibt folglich wesentlich. Die Korrelationen der spezifischen Faktoren korrelieren jeweils mit den spezifischen Faktoren aus anderen Testverfahren der jeweiligen Ebene. Johnson et al. (2007) folgerten daraus, dass keine Identität zwischen den Konstrukten vorliegt.

Resümierend ist zu sagen, dass zahlreiche Hinweise nahelegen, dass Faktoren der Intelligenz von schulspezifisch erforderlichen Fähigkeiten konzeptionell unterschieden werden können. In Bezug auf die Mathematikkompetenz betonen die Autoren, die keine Identität der Konstrukte annehmen, dass Mathematikkompetenz auch mathematikspezifische Fähigkeiten verlangt. Mathematik sei weit mehr als nur das numerische Abarbeiten von Algorithmen, eine Fähigkeit, die vorwiegend intelligenzunabhängig in der Schule erworben wird (Brunner, 2006).

Das Resümee von Baumert et al. (2007) zu der differenzierten Erfassung von Schulleistungen und kognitiven Fähigkeiten sieht keine Identität zwischen Schulleistungen und kognitiven Fähigkeiten, da Schulleistungen Resultate kumulativer Wissenserwerbsprozesse darstellen, die auch durch außerschulische Faktoren und schlussfolgerndes Denken moderiert

werden, sondern eher eine Überlappung oder Unabhängigkeit der Konstrukte. Abschließend konzentrieren sich die Diskussionen mehr und mehr auf die Frage, inwieweit auch kognitive Fähigkeiten ausreichend repräsentiert und nicht noch genauer als nur in fluide und kristallisierte Fähigkeiten unterschieden werden sollen. Insbesondere werden weitere spezifische Fähigkeiten von Kan, Kievit, Dolan und van der Maas (2011) diskutiert.

2.2.1.3 Faktorielle Differenziertheit – Zusammenfassung, Ableitung und Hypothesen

Zusammenfassend muss nun festgehalten werden, dass die eingangs des Kapitels gestellte Frage nicht ohne Weiteres beantwortet werden kann. Für alle theoretisch denkbaren Positionen – Identität, Überlappung und auch Unabhängigkeit – gibt es zahlreiche Verfechter, die ihre jeweilige Position mit ausführlichen Studien untermauern. Den Verfechtern der Identitätstheorie geben vorwiegend die hohen Korrelationen Anlass zu der begründeten Annahme, dass kognitive Fähigkeiten und schulische Leistungen nicht zu unterscheidende Konstrukte sind, zumal die Aufgabenähnlichkeit in den entsprechenden kognitiven bzw. leistungsspezifischen Tests die Unterscheidung grundsätzlich schwierig macht. Auch die ähnlichen Definitionen beider Konstrukte und die implizit vorausgesetzte Kausalität unterstützen ihre Annahme. Kritiker dieser Position sind hingegen von der Überlappung bzw. Unabhängigkeit der Konstrukte überzeugt. In Bezug auf die Frage der Definitionen kritisieren sie das eingeschränkte Verständnis von g, das nicht alle bereichsspezifischen Komponenten der jeweiligen Konstrukte abbildet.

Sie argumentieren außerdem, dass moderierende Faktoren, die die Konstrukte beeinflussen, unberücksichtigt bleiben. Auch die implizit angenommene Kausalinterpretation von kognitiven Fähigkeiten und schulischen Leistungen als identische Konstrukte stößt aufgrund der mangelnden empirischen Beweise auf Widerspruch. Kritik, die die Vertreter der unterschiedlichen Positionen sich gegenseitig zum Vorwurf machen, trifft häufig die Konzeption und das Analysedesign der Studien. Hier wird eine mangelnde Belastbarkeit der Ergebnisse ursächlich konstatiert. Dieser Vorwurf allerdings lässt sich aus Sicht der vorliegenden Arbeit durchaus beiden Seiten machen. Denn den Vertretern beider Positionen sind einige zu kritisierende methodische Unzulänglichkeiten gemein. Deshalb soll nun ein kurzer Blick auf die messtheoretischen Probleme geworfen werden.

Bereits Cattell (1971) und Horn (1989) war bewusst, dass psychometrische Probleme die Aussagekraft dieser Studien stark einschränken können. Häufig basieren bisherige Analysen auf unscharfen, begrifflichen Definitionen (Brunner et al., 2007; Kaufman et

al., 2012). Johnson et al. (2004) mahnen vor der Interpretation hoher Korrelationen als Beweis für identische Konstrukte. Durch die Art und Weise der Auswahl der einzelnen Tests können psychometrische Stichprobenfehler resultieren, die auf der Vermischung der Stratumebenen nach Carroll (1993) basieren, sodass diese über- bzw. unterrepräsentiert werden. Auch Floyd empfiehlt die Beachtung von Messfehlereinflüssen im Zuge seiner Studie von 2013, bei der er Archivdaten anstelle einer neuen Datensammlung verwendete. Das hat zur Konsequenz, dass Unzulänglichkeiten der damaligen Datenerhebungen nicht bekannt waren, aber dennoch die Ergebnisse der Studie beeinflusst haben können. Weiterhin war seine Stichprobengröße insgesamt recht klein, wie auch in einer Folgestudie aus dem Jahr 2013. Allerdings war Floyds Studie methodisch anspruchsvoll konzipiert und eine der ersten, in der Invarianzfaktoren geprüft wurden, um Messfehlereinflüsse zu minimieren.

Kaufman et al. weisen kritisch darauf hin, dass Forscher, die Bildungsleistungsmaße und kognitive Fähigkeiten (COG-g und ACH) als vollkommen korreliert und damit identisch betrachten, häufig Testverfahren einsetzen, die allgemeine schulische Leistungen messen. Außerdem wenden sie ein, dass kognitive Fähigkeiten aus verschiedenen Intelligenztests gemeinsam analysiert werden und damit letztlich ohnehin eine allgemeine kognitive Fähigkeit g als Produkt entsteht. Auch van der Maas et al. (2006) nehmen auf die interagierenden Faktoren Bezug. Wenn zahlreiche Subtests verschiedener Bereiche zum Einsatz kommen, würde in der Summierung ein g-Faktor quasi unweigerlich resultieren. Van der Maas et al. (2006) zeigten in ihrem dynamischen Intelligenzmodell, dass eine dominante latente Variable (g-Faktor) zwar oft mit Leistung assoziiert wird, jedoch dabei die Gegenseitigkeit und positive Wechselwirkungen oftmals unterschätzt werden. Van der Maas et al. (2006) nehmen an, dass verschiedene Aufgaben in Testverfahren letztlich in der Kombination immer einen g-Faktor erzeugen. Kaufmann et al. konstatieren außerdem, dass im Zusammenhang mit dem entsprechenden Konzeptualisierungsverständnis kristallisierte Intelligenz, wie von Cattell seinerzeit (1941) vermutet, eine rein statistische Einheit sei und aus konstruierten Variablen gemessen wird, während die allgemeine kognitive Fähigkeit (COG-g) eher eine latente Variable ist. Daher werden Bildungsleistungsmaße (ACH-g) manchmal mit "kristallisierter Intelligenz" gleichgesetzt, denn sie fasst die Kovarianzen zwischen verschiedenen kristallisierten Fähigkeiten zusammen und kann somit besser als eine zusammengesetzte Variable konzipiert werden (Kan, Kievit, Dolan & van der Maas, 2011 in Kaufman, 2012). Die Schlussfolgerung ist, dass Gc nicht eine einzigartige Kapazität darstellt.

Analysen, die auf der Nutzung nicht kompatibler Testinstrumente beruhen, sind also als problematisch zu betrachten. Ein Versuch, dieser Problematik entgegenzutreten, besteht

in der Nutzung von theoriebasierten Messinstrumenten, die auf einem hierarchisch organisierten Intelligenzmodell fußen. Verschiedene Intelligenztheorien, am bekanntesten Carrolls Drei-Stratum-Theorie der kognitiven Fähigkeiten, Gardners Theorie der multiplen Intelligenz, Cattells Modell fluider und kristallisierter Intelligenz Theorie und auch Sternbergs hierarchisches Intelligenzmodell bieten sich dafür an, denn sie gehen davon aus, dass kognitive Fähigkeiten multidifferenziell zu verstehen sind und interagierende Beziehungen zwischen ihnen bestehen. Auch die CHC-Theorie gehört dazu und ist wichtigste Grundlage für nahezu alle individuellen kognitiven Fähigkeitstests (Kaufman, De Young Gray, Brown & Mackintosh, 2009). In diesen Forschungskreisen besteht Einigkeit darüber, dass die Differenzierung zwischen allgemeiner kognitiver Fähigkeit *g* und spezifischen Fähigkeiten, sowie Lese- und Mathematikleistungen (Flanagan, Keith, McGrew & Vanderwood, 1996; McGrew, Vanderwood, Flanagan & Keith, 1996) wesentlich ist. Die Erfassung der Fähigkeiten auf der zweiten Ebene (fluide Denkfähigkeiten – kristallisierte Denkfähigkeiten) ist darüber hinaus für die pädagogische Praxis wichtig. Sie geben Pädagogen in der Praxis hinreichende Erkenntnisse und reduzieren andere aufwändige Erhebungen. Weiter ist es möglich, Fragestellungen spezifisch zu diagnostizieren und gezielte Interventionen abzuleiten, als wenn nur eine allgemeine kognitive Fähigkeit *g* gemessen werden kann.

Schon Reynolds, Keith, Flanagan, Alfonso (2013) fanden Belege für die Gültigkeit der Cattell-Horn-Carroll-Taxonomie (CHC). Sie verknüpften fünf verschiedene Testbatterien und führten Gültigkeitsstudien durch. Sie konnten anhand konfirmatorischer Faktoranalysen nachweisen, dass das CHC-Modell weiterhin aktuell und richtungsweisend ist. Lediglich der auf der ersten Ordnung der fluiden Denkfähigkeiten (Novel Reasoning) (*Gf*) Faktor war äquivalent mit dem Faktor zweiter Ordnung, sodass von teilweiser Überlappung ausgegangen werden kann (Reynolds, Keith, Flanagan & Alfonso, 2013). Auch Reynolds et al. (2013) konstruierten Modelle, die auf der Cattell-Horn-Carroll-Theorie (CHC) basieren. Sie verglichen verschiedene Intelligenztestverfahren miteinander und nahmen eine Faktorkorrelation höher als .95 an. Die Faktorkorrelationen der in den verschiedenen Intelligenztests ermittelten Werte korrelierten jedoch nicht so hoch wie angenommen. Folglich werde in den verschiedenen Intelligenztests nicht immer genau das Gleiche gemessen. Die Autoren folgern, es sollten zweitens auch spezifische Fähigkeiten wie Verständnis, Wissen, Kurzzeitgedächtnis und Verarbeitungsgeschwindigkeit gemessen werden. Drittens sollten Ergebnisse, die auf verschiedenen Ebenen ermittelt werden, bei der Interpretation der Ergebnisse berücksichtigt werden, denn hier könnten Ungenauigkeiten Quelle für Messfehler

sein. Die Summierung von spezifischen und allgemeinen Fähigkeiten und die daraus ermittelten Gesamtwerte produzieren Interpretationsschwierigkeiten und sind somit nicht zulässig. Vielmehr sollte der Einfluss spezifischer Fähigkeiten auch als solcher interpretiert werden. Reynolds et al. (2013) empfehlen die Testentwicklung, die zum einen allgemeine kognitive Fähigkeiten im Sinne von *g* und darüber hinaus auch spezifische Fähigkeiten erfassen können. Auch für Kaufmans 2012 durchgeführte Studie erscheint das CHC-Modell als theoretisches Modell geeignet, da zehn Fähigkeiten identifiziert werden, wie beispielsweise kristallisiertes Wissen (*Gc*), fluide Denkfähigkeiten (*Gf*), Kurzzeitgedächtnis (*Gsm*) und Prozessgeschwindigkeit (*Gs*). Von den zehn Fähigkeiten lassen sich acht als kognitive und zwei als akademische Leistungsmaße identifizieren (Lesen und Schreiben), darunter quantitatives Wissen. Kaufman et al. (2012) untersuchten die Beziehung zwischen allgemeinen kognitiven Fähigkeitstests, die sich auf die allgemeine kognitive Fähigkeit (COG-*g*) beziehen und denen, die Lese-, Mathematik- und Schreibleistungen erfassen (ACH-*g*). Sie untersuchten in zwei großen nationalen repräsentativen Datensätzen zwei unabhängige Testbatterien und analysierten sie mit konfirmatorischen Faktoranalysen. Es stand ihnen eine Stichprobe von 2520 Kindern im Alter zwischen 4 und 5 und 16 bis 19 Jahren zur Verfügung, die kognitive und Leistungstestbatterien absolvierten. Kaufmann selbst nennt einige Limitationen seiner Untersuchung. Beispielsweise wurden die Bildungsleistungsmaße aus einem Verbund aus Lese-, Schreib- und Mathematikleistungen konzeptualisiert. So ist das Leistungsmaß (ACH-*g*) also als eine gemeinsame Ursache der Korrelationen zwischen diesen akademischen Bereichen zu verstehen. Zuverlässigere Resultate sind zu erzielen, so Kaufman (2012), wenn für ACH-*g*, die Beziehungen aus den Subtests spezifiziert werden. So empfehlen Kaufman et al. zur Darstellung der Beziehungen zwischen Leistungsvariablen und kognitiven Fähigkeiten über die Regressionen aus den Subtests zu favorisieren, da sie präzisere Analysen erlauben, während allgemeine Bildungsmaße einen Summenwert aus der Kombination verschiedener akademischer Leistungsbereiche darstellt und damit keine exakten Ergebnisse liefern kann. Im Anschluss an die Invarianzprüfung, die feststellt, ob kognitive Fähigkeiten und Schulleistungsfähigkeiten im Längsschnitt das gleiche Konstrukt erfassen, wurden die Korrelationen ermittelt. Ein Gesamtdurchschnittskorrelationskoeffizient von .83 konnte ermittelt werden. Im Detail beliefen sich die Korrelationen im jüngeren Alter von .77 bis zu .94 für Kinder in höherem Alter. Die Ergebnisse zeigten, dass die Analysen mittels latenter Faktormodelle keine Bedeutungsgleichheit zwischen kognitiven generellen Faktoren (COG-*g*) und Leistungsfaktoren (ACH-*g*) zeigten, aber die Korrelationen über alle waren mit .83 substantiell, weshalb von einer ähnlichen Beziehung auszugehen ist.

Kaufman warnt allerdings davor, dies als Beleg für die Identität der Konstrukte anzunehmen. Kaufmann betont darüber hinaus die Notwendigkeit der Erfassung der Konstrukte auf Item- oder Subtestebenen. Interpretationen auf Basis von Gesamtscores seien nur eingeschränkt präzise, da sie auf Gesamtwerten zusammengesetzter, konstruierter Variablen beruhen.

Eingeschränkt wird die Qualität von Kaufmans Ergebnissen durch die Tatsache, dass er nur auf Archivdaten zurückgreifen konnte. Er empfiehlt eine Replikation seiner Ergebnisse. Ziel von Folgeforschungen sollte es sein, die präzise Beziehung zwischen allgemeinen kognitiven Fähigkeiten, die auf Subtestebene erhoben werden, und spezifischen Leistungsfähigkeiten zu untersuchen, um die Unschärfe zwischen beiden Konstrukten aufzuklären.

McGrew et al. (1997) nehmen Bezug auf die Ergebnisse einer Sammlung von Studien, die sich der Beziehung zwischen kognitiven Fähigkeiten, im speziellen g, und den spezifischen Fähigkeiten wie spezifischen Lesefähigkeiten und mathematischen Leistungen widmen. Mittels Strukturgleichungsmodellen konnte die Kriterienorganisation auf den genannten hierarchischen Ebenen hinsichtlich ihrer Konsistenz mit Carrolls (193) Drei-Stratum-Modell der Intelligenz getestet werden. Über alle Analysen wurde die Beziehung zwischen allgemeinen kognitiven Fähigkeiten zu genereller Lesefähigkeit und Mathematikleistung erwartet, aber auch zwischen den spezifischen kognitiven Fähigkeiten und den spezifischen Lese- und Mathematikfähigkeiten wurden signifikante Beziehungen gefunden. Das macht es wichtig, die spezifischen kognitiven Fähigkeiten zu erfassen, um die Entwicklung spezifischer Lese- und Mathematikleistungen zu verstehen. Die Debatten über die faktorielle Differenziertheit zwischen Intelligenz und schulischer Leistung werden aktuell von Flanagan und McGrew (1997), angeführt. Flanagan und McGrew (1997), Flanagan, McGrew und Ortiz (2000), Flanagan und Ortiz (2001) und McGrew und Flanagan (1998) entwickelten einen auf dem Cattell-Horn-Carroll-Modell (CHC) basierenden Cross-Battery Assessment Ansatz (XBA), der theoriebasiert den Anwendern von Testverfahren die Möglichkeit gibt, spezifische Fähigkeiten je nach Fragestellung zu testen (Flanagan, Alfonso & Ortiz, 1997). Damit wird die Qualität der Interpretation von Testergebnissen wesentlich fundierter. Auf diesem Ansatz beruhende Testbatterien ermöglichen somit, je nach Auswahl fundierte spezifische Fähigkeiten und Verarbeitungsdimensionen zu erfassen, die Auskunft über kognitive, akademische und neuropsychologische Konstrukte ermöglichen (Flanagan et al., 2013). Ende der 1990er Jahre versuchte McGrew (1997) diese Unterschiede zwischen dem Cattell-Horn- und Carroll-Modell auf der Basis seiner Forschung zu integrieren.

McGrew schlug eine "integrierte" *Gf-Gc*-Theorie (Flanagan et al. 2000) vor, die als Cattell-Horn-Carroll-Theorie (CHC) der kognitiven Fähigkeiten bekannt wurde. Diese besitzt enorme Bedeutung für die Testentwicklung, in der multidimensionale Modelle der Intelligenz berücksichtigt werden. Sie messen also sowohl kognitive als auch akademische Fähigkeiten. Grundlage ist ein dreistufiges hierarchisches Modell der Struktur der kognitiven Fähigkeiten. Auf der obersten Ebene ist ein genereller Intelligenzfaktor angesiedelt. Drei weitere breite Faktoren, die auch kristallisierte Fähigkeiten genannt werden, sowie fluide und analytische Fähigkeiten und Kurzzeitgedächtnis sind auf der nächsten Ebene dargestellt. Auf der dritten Ebene werden spezifische Faktoren repräsentiert, die mündliches Schlussfolgern, quantitatives Denken und abstraktes visuell-logisches Denkvermögen beinhalten. Ferner findet auch der Arbeitsspeicher Berücksichtigung, da seine Bedeutung für den schulischen Erfolg deutlich wurde.

Die Autoren betonen, dass zukünftige Forschung die Bedeutung spezifischer kognitiver Fähigkeiten bei der Erklärung akademischer Ergebnisse berücksichtigen und damit über eine Interpretation eines allgemeinen *g*-Faktors hinausgehen muss, um tatsächlich für Einzelpersonen angemessene Interventionen abzuleiten und Empfehlungen für das Bildungswesen zu treffen.

Zusammenfassend lässt sich schlussfolgern, dass die Aussagequalität bisher vorliegender Studien aufgrund der oben erläuterten methodischen Unzulänglichkeiten zum Teil eingeschränkt ist. Häufig sind die zugrunde gelegten Stichproben älter und teilweise sehr klein, sodass die berichteten Ergebnisse wenig repräsentativ sind. Bei der Datenhebung werden zudem die zu vergleichenden Faktoren nicht präzise genug definiert, sodass die Resultate zu falschen Schlussfolgerungen führen können. Darüber hinaus werden oft zu wenige Faktoren einbezogen, sodass keine belastbaren Aussagen getroffen werden können. Auch die Korrelationen sind oft nur auf manifester Ebene erhoben worden. Eine vorliegende jüngere Studie von Kaufman (2012) erfasste die Korrelationsbeziehungen zwischen generellen kognitiven Fähigkeiten und Bildungsmaßen zwar erstmals auf latenter Ebene, war jedoch als Querschnittstudie angelegt und kann somit keine Aussage über die Zeit treffen. Insgesamt bleibt der Anteil der verbleibenden nicht aufgeklärten Varianz zwischen kognitiven Fähigkeiten und Schulnoten bzw. Schulleistungen damit recht hoch und soll daher zum Gegenstand der ersten Fragestellung dieser Arbeit gemacht werden. Die erste Fragestellung wird daher wie folgt formuliert:

Erfolgen die Erfassung kognitiver Fähigkeiten und die Erfassung mathematischer Schulleistungen innerhalb der Messzeitpunkte (partiell) unabhängig?

Mit der Beantwortung kann ein präzises und methodisch fundiertes Bild zu den expliziten Zusammenhängen zwischen kognitiven Merkmalen von Kindern im Grundschulalter und deren mathematischen Schulleistungen gegeben werden.

Das Anliegen dieser ersten Fragestellung ist es damit, einen grundlegenden theoretischen Beitrag zur Klärung bisheriger kontroverser theoretischer Annahmen hinsichtlich der Unabhängigkeit der Konstrukte zu leisten.

Für die Subskalen der mathematischen Fähigkeiten wurden zum einen Leistungen des Geometriebereichs, Leistungen des Arithmetikbereichs und Leistungen, die den Fähigkeiten im Sachrechnen zuzuordnen sind, erfasst. Allerdings erlernen die Kinder über den Zeitraum von vier Grundschuljahren zwar mathematische Fähigkeiten, die aufeinander aufbauen, dennoch aber unter anderem von dem Textverständnis und von anderen Entwicklungsvorgängen beeinflusst werden, sodass die partielle Unabhängigkeit der Subskalen insgesamt geringer ausfallen sollte als in den Subskalen der KFT-Dimensionen. Es wird eine partielle Unabhängigkeit angenommen.

Aufgrund der im theoretischen Teil begründeten Annahme der Bereichsspezifität kognitiver Leistungen, wird für die Subskalen der KFT-Reihe partielle Unabhängigkeit angenommen, insofern, als dass gewährleistet werden kann, dass verbale Fähigkeiten explizit von quantitativen Fähigkeiten und nonverbalen Fähigkeiten abgrenzbar sind und sich somit eindeutig von den im Mathematikleistungstest erfassten Skalen abgrenzen lassen.

Es wird davon ausgegangen, dass kognitive Leistungen und mathematische Schulleistungen zwar miteinander korrelieren, aber dennoch der Nachweis gelingt, dass unabhängige Konstrukte erfasst werden.

Hypothese 4a:

Es wird erwartet, dass die Konstrukte kognitive Fähigkeiten und mathematische Schulleistungen miteinander korrelieren.

Hypothese 4b:

Es wird erwartet, dass die Konstrukte kognitive Fähigkeiten und mathematische Schulleistungen partielle faktorielle Differenziertheit aufweisen und somit als unabhängige Konstrukte betrachtet werden können.

2.2.2 Fragestellung 2 – Invarianz

Das Hauptanliegen der vorliegenden Arbeit besteht darin, noch ungeklärte komplexe Zusammenhänge zwischen den Konstrukten kognitive Fähigkeiten und mathematische Schulleistungen im Grundschulalter aufzudecken. Zunächst ist dafür die Frage zu klären, inwieweit kognitive Fähigkeiten und mathematische Schulleistungen über die Zeit vergleichbar sind und zweitens, ob die Erfassung beider Konstrukte innerhalb der Zeitpunkte partiell unabhängig von situationsspezifischen Einflüssen und anderen Messfehlern erfolgt. Das ist jedoch nur mit Hilfe von Längsschnittstudien möglich, deren Grundidee darin besteht, präzise Aussagen aus wiederholten Messungen abzuleiten. Die Zeit ist hierbei ein konstituierendes Merkmal, über das Unterschiede im Entwicklungsverlauf, im Sinne von Varianzveränderungen, ermittelbar sind. Längsschnittsdesigns werden daher häufig sogar als unabdingbare Voraussetzung für die Beurteilung der Änderung von Faktoren, die Veränderungen über das chronologische Alter erlauben sollen, betrachtet (Hildebrandt, Robitsch & Wilhelm, 2009). Die Analyse von Längsschnittdaten bietet somit deutliche Vorteile, ist aber auch mit Problemen verbunden. Zum einen müssen die vermutete Ursache und der Effekt tatsächlich miteinander verbunden sein, die angenommene Ursache muss dem Effekt zeitlich vorausgehen und andere Erklärungen müssen für den beobachteten Effekt ausgeschlossen werden können (Hildebrandt, Robitsch & Wilhelm, 2009). Da in Längsschnittstudien dieselben Personen wiederholt zur Messung der relevanten Konstrukte untersucht werden, interessiert vor allem die Stabilität des gemessenen Merkmals über die Zeit. Verändert sich dieses im Laufe der Jahre – bleibt es also nicht stabil – gilt es zunächst, ein in Längsschnittstudien inhärentes Problem auszuschließen. Denn diese erlauben zwar grundsätzlich Aussagen zu den prädiktiven Beziehungen zwischen den Konstrukten und ermöglichen die Darstellung von wechselseitigen Wirkungen, aber nur wenige Längsschnittstudien berücksichtigen dabei den Einfluss von Kovariaten der Veränderungen. Die schwierigen Vergleichbarkeitsbedingungen von Testleistungen im Längsschnitt resultieren so in einem Invarianzproblem. Weitgehend unbeachtet bleibt beispielsweise häufig der Einfluss von nicht beobachtbaren situationsspezifischen Einflüssen über Einjahres-, Zweijahres- und Dreijahres-Abstände. Auch die Ermittlung von Veränderungen potentieller gegenseitiger Einflussfaktoren von Jahr zu Jahr bzw. über zwei bis drei Jahre hinweg fehlt oft. Der Messinvarianzaspekt der eingesetzten Instrumente ist aber ein wichtiges Kriterium der Beurteilung der Qualität empirischer Studien. Darüber hinaus sind Längsschnittstudien, die sich mit den zeitli-

chen Zusammenhängen zwischen kognitiven Fähigkeiten und mathematischen Schulleistungen im Grundschulalter beschäftigen, mit weiteren Problemen konfrontiert. Zumeist orientieren sich Leistungstests im Grundschulalter an der Alterskohorte analog der Klassenstufe. Sie können damit zur querschnittlichen Analyse des individuellen Leistungsniveaus oder auch für interindividuelle Vergleiche herangezogen werden (Hesse & Latzko, 2011). Auch für den Vergleich mit dem am Rahmenplan orientierten Curriculum sind Schulleistungstests geeignet. Dabei sind Ergebnisse, die über standardisierte Schulleistungstests erfasst werden, valider als Schulnoten, die von anderen, wenig messbaren Moderatoren beeinflusst werden (Sauer, 2001). Abgesehen von der insgesamt raren Befundlage an existierenden Längsschnittstudien zu den zeitlichen Zusammenhängen zwischen kognitiven Fähigkeiten und mathematischen Schulleistungen wird die Aussagekraft vieler bislang vorliegender Ergebnisse zusätzlich eingeschränkt, denn insbesondere bei der Berücksichtigung kognitiver und schulischer Leistungen von Kindern kommt es zu einem messtheoretischen Problem. Idealtypisch sind in der empirischen Forschung die Messinstrumente in allen Erhebungen identisch, um sicher zu gehen, dass das Konstrukt identisch erfasst wird. Damit werden Fehlereinflüsse geringgehalten, die sonst fälschlicherweise auf Veränderungen des gemessenen Merkmals der Person zurückgeführt werden könnten. Vom theoretischen Gesichtspunkt ausgehend, liefert die Verwendung ein und desselben Messinstruments also die präzisesten Ergebnisse, da ausgeschlossen werden kann, dass durch den Wechsel der Testverfahren entstehende Ungenauigkeiten mit in die Ergebnisse einfließen. Doch die strukturelle kognitive Entwicklung von Kindern der ersten bis vierten Klassenstufe unterliegt raschen und starken qualitativen Veränderungen (Piaget, 1969, 1972a, 1976; Hofstätter, 1954; Ziegler, 2004; Schneider, 2007; Rost, 2010). Aufgrund der raschen Entwicklungsverläufe (Piaget, 1972a; Sternberg & Powell, 1983; Hany & Heller, 1991; Hasselhorn & Grube, 1996) müssen die Erhebungsinstrumente an die jeweilige Altersstufe angepasst werden, um das Leistungsniveau weiterhin möglichst präzise zu ermitteln. Sonst erzeugen die bis dahin eingesetzten Messinstrumente Deckeneffekte, wenn die Teilnehmer älter werden (Widaman, 2010). Es müssen daher Testadaptionen für die verschiedenen Altersstufen konzeptualisiert werden, damit die Messinstrumente tatsächlich das entsprechende Leistungsniveau ermitteln können. Solche notwendigen Testadaptionen an die Alterskohorte im Kindesalter werfen dann die Frage der Konstruktkonsistenz über die Zeit auf. Die Konstruktkonsistenz der oben beschriebenen Testverfahren, wenn sie im Längsschnitt eingesetzt werden, ist jedoch durch die erwähnten Adaptionen an das chronologische Alter der Untersuchungsteilnehmer zu hinterfragen.

Bei einem Einsatz im Längsschnitt sollten die Verfahren wegen der besseren Konstruktvalidität außerdem auf einem gleichen Konzeptionsverständnis basieren. Häufig unterliegen die einsetzbaren Versionen aber dennoch einem Testautorenwechsel und das Konstruktverständnis variiert zumindest teilweise. Über die Zeit ergibt sich damit die Gefahr von kumulativen Messfehlern. In der Konsequenz müssen deshalb bisher berichtete Ergebnisse zu den prädiktiven Zusammenhängen von kognitiven Leistungen über die Zeit sowie über die Vergleichbarkeit mathematischer Leistungen im Längsschnitt auf den Prüfstand gestellt werden. Darüber hinaus stand zumeist die Schulleistungsentwicklung im Zusammenhang zum kognitiven Leistungspotential eines früheren Testzeitpunktes im Fokus der wenigen bisher vorliegenden Längsschnittuntersuchungen. So wurde die Leistungsveränderung gegenüber dem vorherigen Messzeitpunkt betrachtet, meist unter Hinzuziehung einer Kontrollgruppe, mit dem Ziel, Einflüsse von Traitmaßnahmen zwischen den Gruppen zu untersuchen. Da die mathematischen Kompetenzen, die im Grundschulalter erworben werden, eine bedeutende Rolle hinsichtlich zukünftiger Schullaufbahnentscheidungen spielen (Hofstätter, 1954; Schneider, 2007; Helmke & Weinert, 1997; Heller, 1991), sind präzise Ergebnisse aber besonders relevant. Für belastbare Ergebnisse ist jedoch eine möglichst hohe Anzahl an Messzeitpunkten erforderlich. Denn nur, wenn die interessierenden Merkmale über die Zeit valide erfasst werden, lassen sich über Entwicklungsverläufe genaue Aussagen formulieren (Köller, 2008). Die Frage nach der Vergleichbarkeit von Messungen über die Zeit nimmt also in Längsschnittstudien eine wesentliche Rolle ein. Von Interesse ist, inwieweit die psychometrischen Eigenschaften der Messinstrumente (Indikatoren) über die Zeit hinweg vergleichbar sind, ob also tatsächliche Veränderungen der Phänomene oder lediglich Veränderungen zwischen latenten Variablen gemessen werden. Weiterhin muss sichergestellt werden, dass die gemessenen Leistungen durch vergleichbare Konstrukte erklärt werden können (Geiser, 2011). Die zentrale Frage nach der Vergleichbarkeit kann mit der Testung der Messinvarianz beantwortet werden. Mit der Testung der Messinvarianz werden folglich die Indikatoren (beobachtete Variablen) mit den latenten Variablen (nicht beobachteten Variablen) verknüpft und auf ihre Vergleichbarkeit hin getestet. Wenn sich keine Invarianz in der Datenstruktur nachweisen lässt, sind Gruppenvergleiche oder auch Vergleiche über mehrere Testzeitpunkte mit wechselnden Verfahren nicht möglich und die Messinstrumente sind für den gewünschten Vergleich nicht zu empfehlen bzw. bedürfen einer Revision. Die Berücksichtigung der Invarianzbedingungen der eingesetzten Testverfahren ist also für die Aussagequalität der Ergebnisse entscheidend. Im Folgenden sollen daher bislang vorliegende Längsschnittstudien, die für das Thema der Arbeit relevant sind, unter

dem Aspekt betrachtet werden, inwieweit Messfehler und Invarianzbedingungen berücksichtigt wurden.

Die Stabilität des gemessenen Merkmals und die Ursachen möglicher Veränderungen über die Zeit standen bislang nur in wenigen Studien im Mittelpunkt des Interesses (Übersicht bei Watkins, 2007, Ceci, 1991, Soares, 2015). Explizit für das Grundschulalter konnte nur eine Studie mit zwei Messzeitpunkten (Watkins, 2007) recherchiert werden.

Einen der ersten Versuche, die Beziehungen zwischen beiden Konstrukten präzise zu analysieren, unternahmen Watkins, Lei und Canivez (2007). Sie widmeten sich damit der Thematik, inwiefern kognitive Fähigkeiten bzw. schulische Leistungen dominierenden Einfluss auf das jeweils andere Konstrukt haben. Sie konstruierten ein Modell über zwei Messzeitpunkte und untersuchten die Pfade von Leistungsergebnissen zu Bildungsleistungen in gegenseitiger Richtung. Die Studie umfasste zwar nur zwei Messzeitpunkte, wurde aber auf hohem methodischem Niveau mittels der Spezifikation von Strukturgleichungsmodellen sowie unter Prüfung der Invarianzbedingungen analysiert. Ausgangsgehend von einer Stichprobe von 289 Kindern im Alter von durchschnittlich 9,25 Jahren (6,0 bis 13,9 Jahre) wurde die Studie mit zwei Messzeitpunkten konzipiert. Nach 2,8 Jahren waren die Kinder durchschnittlich 12,08 Jahre (8,0 bis 16,9 Jahre). Die Ergebnisse ihrer Studie ermittelten einen Einfluss kognitiver Fähigkeiten auf die Bildungsleistungen, während umgekehrt Leistungsergebnisse zukünftige Intelligenz Leistungen nicht wesentlich beeinflussen. Leider basieren die Ergebnisse nur auf einer vorausgelesenen, recht kleinen Stichprobe von Kindern mit sonderpädagogischen Fragestellungen, die daher nur eingeschränkt repräsentativ ist, so Watkins selbst. Zudem konnten nur wenige Kinder innerhalb von zwei Jahren erneut getestet werden, da einige nicht mehr die Sonderschulen besuchten. Watkins et al. (2007) haben die Invarianz über die Zeit für die Untertests des eingesetzten kognitiven Fähigkeitstests geprüft. Sie untersuchten die Invarianz im Zusammenhang mit der oben genannten Fragestellung. Diese Fragestellung wird im Rahmen der vorliegenden Arbeit in der vierten Fragestellung diskutiert. Watkins et al. (2007) wollten zudem beantworten, inwieweit schulische Leistungskonstrukte identisch mit kognitiven Fähigkeiten sind. Diese Fragestellung der Autoren wird im Rahmen der vorliegenden Arbeit in Frage 4 aufgenommen und diskutiert. Für die Erfassung kognitiver Fähigkeiten setzten Watkins et al. (2007) den Wechsler-Intelligenztest für Kinder (Wechsler WISC – III) ein. Dieser Test erfasst bereichsspezifische Fähigkeiten in 13 Subtests, von denen zwei zur Anwendung kamen (verbale Fähigkeiten und Wahrnehmungsorganisation). Die akademischen Leistungen wurden mit 5 ausgewählten Kombinationen des Woodcock-Johnson Tests of Achievement, dem

Wechsler Individual Achievement Test und dem Kaufman Test of Educational Achievement erfasst, mit denen die Leseleistungen (word reading), die mathematischen Rechenfertigkeiten (separate calculation) und das schlussfolgernde Denken (reasoning subtests). Die Passung des CFA-Modells lässt, so Watkins et al. eine wahrscheinliche Invarianz vermuten, da die Faktorladungen zu beiden Zeitpunkten ähnlich ausgeprägt waren. Es ist den Ergebnisberichten jedoch nicht zu entnehmen, ob die Leistungstests ebenfalls invariant über die Zeit waren. Das Problem der Auswirkungen fehlender Invarianzprüfung thematisierten auch Kaufman et al. (2012). Die Autoren untersuchten die Beziehung zwischen allgemeinen kognitiven Fähigkeitstests, die sich auf die allgemeine kognitive Fähigkeit (COG-g) beziehen und denen, die Lese-, Mathematik- und Schreibleistungen erfassen (ACH-g). Sie untersuchten in zwei großen nationalen repräsentativen Datensätzen zwei unabhängige Testbatterien und analysierten sie mit konfirmatorischen Faktoranalysen. Es stand ihnen eine Stichprobe von 2520 Kindern im Alter zwischen 4 und 5 und 16-19 Jahren zur Verfügung, die kognitive Tests und Leistungstestbatterien absolvierten. Im Anschluss an die Invarianzprüfung, die feststellt, ob kognitive Fähigkeiten und Schulleistungsfähigkeiten im Längsschnitt das gleiche Konstrukt erfassen, wurden die Korrelationen ermittelt. Es konnte ein Gesamtdurchschnittskorrelationskoeffizient von .83 ermittelt werden. Im Detail beliefen sich die Korrelationen im jüngeren Alter von .77 bis zu .94 für Kinder in höherem Alter. Diese Ergebnisse widersprechen aber früheren Ergebnissen von Jensen (1998), der ein Absinken der Korrelationsbeziehung zwischen Intelligenz und Bildungsleistung mit zunehmendem Schulalter empirisch zeigte. Kaufman et al. (2012) bemerken, dass in den Ergebnissen von 1998 die fehlende Invarianzprüfung ursächlich gewesen sein könnte. In ihrer Studie von 2012 konnten Kaufmann et al. gültige Vergleiche über das Alter hinweg aufgrund der Invarianzprüfung nachweisen. Das Problem fehlender Invarianzbedingungen machten auch Floyd et al. (2013) zum Thema. Sie postulierten und prüften ein Modell, das speziell intellektuelle Fähigkeiten von Kindern und Jugendlichen beschreibt. Es basiert auf der CHC-Faktortheorie, nimmt aber auf zwei Ebenen allgemeine und spezifische Intelligenzfaktoren an (Floyd, 2013). Ausgangspunkt der Forschungsbemühungen war die Feststellung, so Floyd, dass bis dato nur zwei Zeitschriftenartikel existieren, die sich der Korrelationsbestimmung verschiedener Intelligenztestverfahren im Kindesalter widmeten. Seine Stichprobe setzte sich aus 200 Kindern im Alter von 6 bis 17 Jahren sowie weiteren 135 Kindern im Alter von 8 bis 13 Jahren zusammen, die verschiedene kognitive Testverfahren absolvierten. Insgesamt basierte die Stichprobengrundlage auf Daten von 486 Kindern, die mehrere Intelligenztests absolvierten. Floyd et al (2013) stellten sich die Frage, ob der in den

verschiedenen kognitiven Fähigkeitstests ermittelte g-Faktor über die unterschiedlichen Verfahren hinweg das Gleiche misst. Zudem untersuchten sie die spezifischen Fähigkeiten, die in den Tests erhoben wurden und ermittelten die Korrelationen. Damit stellten sich Floyd et al. dem Problem der Invarianz kognitiver Fähigkeitstests, um Messfehlereinflüsse zu minimieren. Sie beabsichtigten vor allem den Argumenten von Jensen (1998) zu begegnen, der die Interpretierbarkeit von Ergebnissen durch psychometrische Stichprobenfehler eingeschränkt sah. Diese entstehen immer dann, wenn kognitive Aufgaben verschiedener hierarchischer Ebenen gemischt, also nicht hierarchisch erhoben werden. So plädiert er für die bereichsspezifische Erhebung von z. B. Gedächtnisaspekten und verbalen oder räumlichen Inhalten. Ein wesentliches Ergebnis Floyds et al. war die Bestätigung, dass der allgemeine Intelligenzfaktor als Faktor höherer Ordnung modelliert werden konnte. Die Korrelationen hatten eine Höhe von .89 bis .92. Aufgrund der relativ kleinen Stichprobenumfänge empfehlen aber Floyd et al. (2013), die Ergebnisse zu replizieren und ferner keine Archivdaten (wie in ihrer Untersuchung) als Datengrundlage zu nehmen. Damit belegen die Ergebnisse theoriekonforme Modellierung eines Intelligenzfaktors auf höherer Ebene, wie beispielsweise von Keith, Kranzler und Flanagan (2001), basierend auf der Cattell-Horn-Carroll (CHC) Theorie (Schneider & McGrew, 2012), die die Struktur der Intelligenz ebenso hierarchisch modellierte. Auch die Ergebnisse von Floyd et al. (2010) zeigten nahezu perfekte Korrelationen zwischen den allgemeinen Intelligenzfaktoren zweier untersuchter Testverfahren. Für die Subtestvergleiche ließ sich ebenso eine recht hohe Konsistenz der Subtestladungen zeigen. Es sei an dieser Stelle erwähnt, dass sich Floyd et al. (2013) gegen ein identisches Konstruktverständnis der zugrundeliegenden Tests aussprechen, da sich auch in der Messung der spezifischen kognitiven Fähigkeiten Konstruktvaliditätsprobleme zeigten. Diesen könne mit Strukturgleichungsmodellen begegnet werden. Vorteil neuerer Strukturgleichungsmodelle ist die Möglichkeit zur expliziten Differenzierung zwischen latenten Bestandteilen eines Konstrukts und deren manifester Indikatoren. Insofern können Messfehler explizit extrahiert (Perels, Otto & Schmitz, 2008) und Merkmalszusammenhänge exakter erklärt werden. Ihren Ursprung haben Strukturmodelle, sogenannte SEM, in der Pfadanalyse. Sie gehen zwar als generalisierte (multivariate) Regressionsmodelle auf die Pfadanalyse und die konfirmatorische Faktorenanalyse zurück (Bortz, 2004), besitzen jedoch einen wesentlichen Vorteil, da sie unbeobachtete Variablen aufnehmen können (Eid et al. 2013; Geiser, 2011). Ein Strukturmodell setzt sich aus der Anzahl der latenten Variablen mit den zugehörigen Parametern eines einfachen Messmodells zusammen und enthält zusätzlich mindestens eine Kovarianz/Korrelation.

2.2.2.1 Invarianz – Zusammenfassung, Ableitung und Hypothesen

In dieser Arbeit stehen die prädiktiven und gegenseitigen Zusammenhänge zwischen kognitiven Fähigkeiten und Schulleistungen im Grundschulalter im Mittelpunkt. Es soll überprüft werden, inwiefern kognitive Fähigkeiten und Mathematikleistungen über einen Zeitraum von vier Messzeitpunkten im jährlichen Abstand überhaupt vergleichbar sind, denn die Sichtung entsprechender Untersuchungen zeigt, dass die eingesetzten Testverfahren zwar den Gütestandards für querschnittliche Untersuchungen hinreichend genügen, jedoch nur selten Auskunft darüber geben, inwieweit präzise die Messinstrumente einem längsschnittlichen Studiendesign Rechnung tragen.

Um nachfolgend genaue Aussagen zu formulieren, soll zunächst die grundlegende Frage, wie zuverlässig Testverfahren die interessierenden Konstrukte im Längsschnitt messen können, beantwortet werden.

Daher widmet sich die zweite Fragestellung dem Problem der Vergleichbarkeit der Konstrukte kognitiver Fähigkeiten und Mathematikleistungen von Grundschulkindern über einen Zeitraum von vier Jahren. Ob sich diese Invarianz in den in dieser Arbeit eingesetzten Testverfahren nachweisen lässt, führt zur Formulierung der Fragestellung 2. Die zweite Forschungsfrage betrachtet somit den Aspekt der Messfehler und ihrer Ausprägung und lautet:

Ist die Messung kognitiver Fähigkeiten und Mathematikleistungen vergleichbar über die Zeit?

Kognitive Fähigkeiten

Zur zweiten Fragestellung werden folgende Hypothesen formuliert: Das Testverfahren zur Messung kognitiver Fähigkeiten (KFT) liefert in der Konzeption für die erste und zweite Klassenstufe, in der Konzeption für die dritte Klassenstufe und in der Konzeption für die vierte bis zwölfte Klassenstufe valide und reliable Abbildungen des Konstrukts der Intelligenz (Abou-Koura & Perleth, 2005; Perleth & Heller 2008; Heller & Perleth, 2000). Daher wird trotz, oder eben gerade wegen der Adaption der Testversionen an die jeweilige Intelligenzentwicklungsstufe von Kindern im Grundschulalter eine hohe Vergleichbarkeit der kognitiven Fähigkeitstests über die vier Messzeitpunkte im Jahresabstand erwartet. Da wegen der Anpassung an das chronologische Alter der Kinder Einschränkungen zu erwarten sind, wird von einer gewissen Beeinträchtigung der Konstruktvalidität über die vier Jahre ausgegangen, die sich über latente Variablen ermitteln lässt.

Die Hypothese wird wie folgt formuliert:

Hypothese 2a:

Es wird erwartet, dass Invarianz der Verfahren zur Erfassung kognitiver Fähigkeiten der KFT-Reihe für die vier Messzeitpunkte im jährlichen Abstand vorliegt.

Mathematische Schulleistungen

Auch für dieses Konstrukt wird davon ausgegangen, dass die Vergleichbarkeit mathematischer Fähigkeiten über die vier Messzeitpunkte hoch ist. Dabei wird dennoch erwartet, dass latente, also unbeobachtete Einflüsse, die Konstruktvalidität beeinträchtigen.

Daher wird die Hypothese wie folgt formuliert:

Hypothese 2b:

Es wird erwartet, dass Invarianz der Verfahren zur Erfassung mathematischer Schulleistungen der DEMAT-Reihe für die vier Messzeitpunkte im jährlichen Abstand gegeben ist.

2.2.3 Fragestellung 3 – Autoregressive Effekte kognitiver Fähigkeiten und mathematischer Schulleistungen

Intelligenz gilt allgemein als ein sehr stabiles Merkmal, welches entwicklungspsychologisch gesehen gegen Ende des Grundschulalters nur noch relativ geringen Veränderungen unterliegt. Auch das Niveau mathematischer Schulleistungen wird als relativ stabil betrachtet. Werden Intelligenztestverfahren eingesetzt, so wird davon ausgegangen, dass das einmal festgestellte Niveau kognitiver Leistung einer Person auch relativ reliabel mit dem Ergebnis eines Testverfahrens zu einem späteren Zeitpunkt übereinstimmen würde. Daher werden Intelligenztestverfahren nur selten wiederholt eingesetzt, sondern es wird vornehmlich davon ausgegangen, dass eine einmal festgestellte kognitive Fähigkeit den Erfolg bei der Auseinandersetzung verschiedenster Anforderungen zu (beliebig) vielen Zeitpunkten vorhersagen könne. Der Zugang zu bestimmten Förderprogrammen oder zur Vorhersage von Studien- oder Berufserfolg beruht daher oftmals auf einer einmalig festgestellten kognitiven Fähigkeit. Theoretischer Hintergrund ist dabei die Annahme, dass frühere kognitive Fähigkeiten als verhältnismäßig stabil über die Zeit betrachtet werden und damit relativ unabhängig von anderen Faktoren niveaustabil über die Zeit bleiben. Im folgenden Kapitel soll nun darüber Aufschluss gegeben werden, wie in der bisherigen Forschungslage die Frage nach der Entwicklung und der Stabilität kognitiver Fähigkeiten und mathematischer Schulleistungen über einen längeren Zeitraum im Grundschulalter beantwortet wurde. Die

Antwort auf die Frage nach der Stabilität der Konstrukte im Grundschulalter ist von großem Interesse im pädagogisch-psychologischen Kontext, denn auf ihr fußt das Bemühen, den Schulerfolg mittels Förderung und Forderung zu optimieren.

Im Kontext der Frage ist deshalb zunächst eine Definition des Begriffs Stabilität, also der Unveränderlichkeit von Merkmalen über die Zeit, erforderlich. Während bei absoluter Stabilität über die Zeit keine Veränderung gemessen wird (Schneider & Lindenberger, 2012), bleibt bei der strukturellen Stabilität der Zusammenhang zwischen den einzelnen Merkmalen über die Zeit konstant. Der Mittelwert einer Gruppe ist in diesem Fall über die Zeit stabil, es können aber trotzdem individuelle Veränderungen einzelner Personen vorliegen. Bei der intraindividuellen Stabilität werden individuelle Veränderungen über die Zeit erfasst. Dabei stehen die Veränderungen des Niveaus innerhalb einer Person im Zentrum der Fragestellung. Mit der Erfassung wird beantwortet, ob kognitive Fähigkeiten der Person im Niveau zu- oder abgenommen bzw. unverändert geblieben sind (Rost, 2009). Im Gegensatz dazu werden bei der Fokussierung auf *interindividuelle* Stabilität die Veränderungen von Personen im Vergleich zu der Veränderung anderer Personen in das Zentrum der Forschung gestellt. Diese wird auch als Positionsstabilität bezeichnet und wird immer dann ermittelt, wenn die Position von Personen in einer Rangreihe über die Zeit betrachtet werden soll (Rost, 2009). Nach einem kurzen Blick auf den Forschungsstand zu der strukturellen sowie der intraindividuellen Stabilität im Zusammenhang zwischen kognitiven Fähigkeiten und mathematischen Schulleistungen soll die *interindividuelle* Stabilität der beiden Konstrukte über die Zeit im Zentrum der Betrachtung stehen. Sie ist im Rahmen dieser Arbeit entscheidend.

2.2.3.1 Autoregressive Effekte der kognitiven Entwicklung über die Zeit

Die Strukturstabilität bezieht sich, wie beschrieben, auf die qualitative Veränderung eines gemessenen Konstrukts. Betrachtet man sie in Bezug auf die strukturelle kognitive Entwicklung von Kindern, ist die bereits oben angerissene Beschreibung von Piaget (1936, 1969, 1972a) am bekanntesten. Er machte die qualitative Entwicklung kognitiver Fähigkeiten vom Säuglingsalter bis zum Erwachsenenalter zum Gegenstand seiner Forschungen und entwickelte ein Stufenmodell zu ihrer Beschreibung (Rost, 2009). Er unterschied vier Stufen der kognitiven Entwicklung: die sensomotorische Stufe als erste, auf die als zweite Stufe die präoperationale Stufe folgt, die von der konkret-operationalen Stufe abgelöst wird und in eine vierte, formal-operationalen Stufe mündet, die im Lebensalter von etwa 15 Jahren voll ausgebildet ist (Piaget, 1969). Damit sieht Piaget die kognitive Entwicklung als vollendet

an. Strukturstabilität ist nach diesem Modell dann erreicht, wenn kaum mehr altersbedingte strukturelle Unterschiede zwischen dem Denken jüngerer Kinder zwischen und dem Denken älterer Kinder nachweisbar sind. Diese betreffen die Art und Weise, wie Wissen repräsentiert werden kann, die logischen Denkoperationen, die ausgeführt werden können und das Verständnis fundamentaler Begriffe wie Zeit, Raum, Zahl und Kausalität. So ist das Denken des Vorschulkindes noch gekennzeichnet durch die Unfähigkeit, zwischen Schein und Wirklichkeit zu unterscheiden, durch Egozentrismus, durch die Unfähigkeit, mehrere Merkmalsdimensionen gleichzeitig zu berücksichtigen (Piaget, 1936, 1969, 1972a). Die Literaturlage zeigt, dass die Struktur der Intelligenz bei jüngeren Kinder noch recht instabil zu sein scheint (Ziegler, 2004; Schneider, 2007; Rost, 2010), während sie bereits gegen Ende des Grundschulalters als relativ stabil anzusehen ist (Trost et al., 1996; Franić, 2014).

Obwohl Piagets Modell weitgehende Bekanntheit erlangte, so ist es doch wegen seiner recht starren Stufenorientierung umstritten, da es keinen Spielraum für kulturelle, geschlechtsspezifische oder kulturelle Unterschiede (Anderson, 2007) zulässt. Auch das scheinbare Ende der kognitiven Entwicklung mit dem Eintritt in das Jugendalter wird kritisiert. Aufgrund dieser Defizite in Piagets Modell begann die Entwicklung bereichsspezifischer Theorien kognitiver Entwicklung. Wie oben erwähnt, nimmt die quantitative Veränderung kognitiver Fähigkeiten von Geburt an bis zum Alter von etwa neun Jahren zügig zu (Bloom, 1971). Die allgemeinen kognitiven Fähigkeiten im Sinne von Spearman erfahren eine deutliche Zunahme bereits in der frühen Kindheit. Das lässt sich leicht feststellen, wenn ein Kind mit zunehmendem Alter eine größere Menge an Aufgaben eines Intelligenztests zu bearbeiten vermag (Rost, 2009). Betrachtet man speziell das Grundschulalter, nehmen kognitive Fähigkeiten zunächst stark zu, während sie ab dem Alter von 9 Jahren dann weniger stark zunehmen (Bloom, 1971) und etwa um das 17. Lebensjahr ein individuelles Höchstniveau erreicht wird (Holling et al., 2004b; Oerter, 1987), das ab dem 21. Lebensjahr wieder abnimmt. In diesem Zusammenhang werden die Berkeley-Grow-Study (Baylay, 1949), die Child-Research-Council-Study (Hilden, 1949) und die California-Guidance-Studx (Honzik et al., 1948) als relevante Studien am häufigsten zitiert.

Übereinstimmend gehen verschiedene Theorien kognitiver Entwicklung davon aus, dass das Denken von Kindern sich vom Denken Erwachsener hauptsächlich durch die Geschwindigkeit und Effizienz von Prozessen der Informationsverarbeitung und durch die Kapazität des Arbeitsgedächtnisses unterscheidet. Ältere Kinder können Informationen schneller und effizienter verarbeiten als jüngere Kinder. Kinder unterscheiden sich aber auch durch ihr Ver-

ständnis wichtiger Inhaltsbereiche, z. B. durch ihr Verständnis der physikalischen Objektwelt, mathematischer Operationen und psychologischer Vorgänge. Die Entwicklung des Denkens in der Kindheit ist also beschreibbar als Entwicklung des Verständnisses wichtiger Wissensbereiche. Es lässt sich soweit sagen, dass das Niveau kognitiver Leistungen ab dem frühen Erwachsenenalter eine recht hohe Niveaustabilität erreicht (Ziegler, 2004; Rost, 2010; Schneider, 2007). Für eine annähernd stabile Struktur kognitiver Leistungen bereits ab dem zehnten Lebensjahr sprechen die Befunde von Rost (2010; Trost et al., 1996). Gottfried et al. (2009) ermittelten über ein Zeitintervall von neun Jahren (Alter der Kinder bei der Ersttestung war acht Jahre und bei Abschlusstestung 17 Jahre) Korrelationen in Höhe von .67 bis zu .77. In der von Deary et al. (2000) berichteten umfassenden Längsschnittstudie mit Daten aus dem 1932 Scottish Mental Survey wurden über 87000 Personen im Alter von elf Jahren erstmals und bis zum Alter von 77 Jahren mit noch 97 Personen hinsichtlich ihrer kognitiven Fähigkeiten verfolgt und noch weiterhin untersucht. Die Ergebnisse weisen eine Korrelation über die Lebensspanne von .63 bis .73 und damit eine erhebliche Stabilität auf. Im Jahr 2004 berichteten Deary et al. (2004) über Stabilitätskoeffizienten zwischen den vormals Elfjährigen und nunmehr Achtzigjährigen in Höhe von .66. Deary (2014) schlussfolgert aufgrund seiner Reanalyse mit Daten der schottischen Längsschnittstudie aus den Jahren 1932 bis 1947, worin 50 % der Teilnehmer hohe Stabilitätswerte der kognitiven Fähigkeiten aufwiesen, eine hohe Stabilität über die Lebensspanne.

Es werden jedoch in der differenzierten Betrachtung verschiedene Positionen deutlich. Insbesondere hinsichtlich der weiteren Ausdifferenzierung kognitiver Fähigkeiten herrscht noch Unklarheit. In einer Längsschnittstudie von Schalke et al. (2013) zur Stabilität von Intelligenz über einen Zeitraum von 40 Jahren stellten die Autoren fest, dass sich die kristallisierten Intelligenzanteile im Sinne Cattells auch noch weiter mit zunehmendem Alter ausdifferenzieren, während eine große Stabilität für die sogenannten fluiden Denkfähigkeiten vorgefunden wurde. Diese Position vertritt auch Ziegler (2004), der hinsichtlich der intraindividuellen Stabilität von Intelligenz zunächst die rapide Zunahme kognitiver Fähigkeiten betont und insgesamt von einer niedrigen Stabilität im Kindesalter ausgeht. Ob jedoch eine früh ausgeprägte hohe intellektuelle Ausstattung später weiterhin eine entsprechende Ausprägung hat, hängt vom Zeitpunkt des Beginns der Förderung ab und den anregenden Lernumweltbedingungen ab, die ein Kind vorfindet. Ziegler begründet seine Position mit Ergebnissen der Scholastik- und der Longitudinalstudie zur Genese individueller Kompetenzen (LOGIK). Die LOGIK-Studie hatte zum Ziel, den Verlauf der kindlichen Entwicklung in verschiedenen Entwicklungsbereichen zu erfassen. In den Jahren 1984 bis

2004 wurden 200 Kinder vom 4. bis zum 24. Lebensjahr hinsichtlich ihrer Entwicklungen im Kindesalter umfassend begleitet. In beiden Studien konnte nur eine niedrige bis mittlere Korrelation zwischen den Messungen im 4., im 12. und im 24. Lebensjahr festgestellt werden (Ziegler, 2004). Dennoch schränkt Ziegler die Aussagequalität der berichteten Ergebnisse ein, indem er auf Messfehlereinflüsse verweist, die unter Einsatz verschiedener Testverfahren und dem Einsatz von Analyseverfahren auf korrelativer Basis entstehen können (Ziegler, 2004). Diesem Problem wird in der vorliegenden Arbeit in Frage 2 nachgegangen. Ziegler selbst versuchte, die Messfehlereinflüsse zu minimieren, indem er zwei Intelligenztests kombiniert (Ziegler, 2004). Cattell (1971) und Horn und Cattell (1967) nahmen an, dass nach einem steilen Anstieg kognitiver Fähigkeiten in der frühen Kindheit die weitere Zunahme fluider Intelligenzanteile langsamer voranschreitet und ihr höchstes Ausmaß im frühen Erwachsenenalter annimmt, während die kristallisierte Intelligenz weiterhin kontinuierlich zunimmt (Cattell, 1971; Horn & Catell, 1967). Diese Feststellungen von Cattell und Horn werden jedoch nicht uneingeschränkt geteilt. Flynn beispielsweise (1987) berichtete, dass nicht die kristallisierten Intelligenzanteile stärker zunehmen, sondern stattdessen fluide Denkfähigkeiten. Auch Cahan und Cohen (1989) berichteten die größeren Gewinne für fluide Anteile der Intelligenz (in Ceci & Williams, 1997).

Auch uneinheitlich sind bisher vorliegende Befunde, die zum einen von einem Anstieg kristallisierter Intelligenz über die Lebenszeit ausgehen, während andere Studien sogar gegenteilige Ergebnisse offenbaren und nicht von einer weiteren Ausdifferenzierung ausgehen. Schroeders et al. (2015) fanden beispielsweise für das Kindesalter einen zunehmend geringeren Anstieg fluider Denkfähigkeiten jedoch beobachteten sie auch eine geringe Zunahme kristallisierter Intelligenzanteile. Mit diesen Ergebnissen wird die weitere Differenzierung kristallisierter Denkfähigkeiten im Schulalter infrage stellt. Schroeders et al. (2015) erklären den Befund mit der gemeinsam geteilten Lernumgebung, die die weitere Zunahme und Ausdifferenzierung von Wissensselementen (Schroeders et al., 2015) verhindert. Schalke et al. (2013) untersuchten über eine Zeitspanne von 40 Jahren die Stabilität und Veränderung der Struktur der Intelligenz hinsichtlich der Differenzierung bzw. Dedifferenzierung kognitiver Fähigkeiten anhand der Daten von 344 Personen. Ihre Ergebnisse zeigten, dass sich die Menschen im Laufe der Zeit stärker unterscheiden und es eine erhebliche Ausdifferenzierung im Alter zwischen 12 und 52 Jahren gibt. Insbesondere waren diese Effekte auf die unterschiedliche Entwicklung der allgemeinen kognitiven Fähigkeiten zurückzuführen. Die Autoren diskutieren diese Strukturveränderung gemäß der Investitionstheorie. Die Beobachtungen von Horn und Cattell (1967) zur Differenzierung fluider und

kristallisierter Intelligenzanteile über verschiedene Altersgruppierungen beschreiben ein höheres Niveau fluider Denkfähigkeiten jüngerer Personen (Vierzehn- bis Siebzehnjährige), während ältere Erwachsene ein höheres Niveau kristallisierter Intelligenzanteile besitzen im Vergleich zu den jüngeren Personengruppen. In der Personengruppe 21- bis 28-Jähriger findet sich das höchste Niveau für visuelle Wahrnehmung, das dann wiederum abnimmt. Für Geschwindigkeitsfaktoren konnten keine systematischen Effekte festgestellt werden.

Auch die interindividuelle Stabilität innerhalb von Personengruppen, die hinsichtlich ihrer kognitiven Voraussetzungen unterschieden wurden, wird in Forschungskreisen diskutiert: Im Rahmen der Marburger Hochbegabtenstudie von Rost (1993) wurde eine weniger hohe Stabilität für die Gruppe der Hochbegabten ermittelt (Rost, 1993). Die erneute Klassifikation der Hochbegabung mit einem Cut-off-Wert von 130 in der Intelligenzgesamtleistung im Alter von 15 Jahren nach einer Erstdiagnose im Alter von 9 Jahren zeigte nur eine erneute Hochbegabung bei 50 % der Kinder. Insofern geht Rost von einer mäßig hohen bis geringen Stabilität aus. In einer Reanalyse der Daten aus der LOGIK Studie von Schneider, Niklas und Schmiedeler aus dem Jahr 2014 wurde die interindividuelle Stabilität zwischen der Ersttestung im Alter von 4 bis 12 Jahren mit den Werten der Personen, die im Alter von 17 bis 23 Jahren verglichen. Die Stabilität war insbesondere für die Gruppe der kognitiv Schwächeren sehr hoch. Auch hier ist zu berücksichtigen, dass die Adaptionen der Testverfahren unberücksichtigte Messfehlereinflüsse hinsichtlich der Konstruktkonsistenz aufweisen können. Hanses erwähnt zudem die von Schallberger (1991) berichtete eingeschränkte Zuverlässigkeit älterer Intelligenztests unter dem Aspekt der Normverschiebung (Hanses, 2000). Hinsichtlich der interindividuellen Stabilität kognitiver Fähigkeiten gehen zahlreiche Forscher davon aus, dass die Position, die ein Individuum im Vergleich zu einer Gleichaltrigengruppe einnimmt, sich nur unwesentlich ändert (Oerter, 2002; Ziegler, 2004). Sehr anschaulich lässt sich diese Positionsstabilität mittels des Intelligenzquotienten beschreiben, der die relative Position der individuellen Intelligenz im Gleichaltrigenvergleich angibt (Ziegler, 2004). Der interindividuelle Vergleich zeigt, dass sich Intelligenzunterschiede bereits um das zehnte Lebensjahr auf einem individuellen Niveau zu stabilisieren scheinen (Trost et al., 1998). Dieses individuelle Niveau kognitiver Fähigkeiten und sein weiteres Steigerungspotenzial sind aber stark vom Anregungsgehalt der Umwelt und von nicht kognitiven Persönlichkeitsmerkmalen abhängig (Heller, 2000; Perleth, 1997). Sauer und Gamsjäger (1996a) stellten in ihrer Längsschnittuntersuchung fest, dass die interindividuelle Stabilität mit Beginn des Schuleintritts stark zunimmt und dann bereits im Grundschulalter hoch stabil bleibt. Auch Helmke und Schrader (2010) stellten in der Münchner

SCHOLASTIK Studie eine hohe Positionsstabilität fest. Schneider und Stefanek (2004) führten eine Reanalyse der LOGIK-Studie durch. Auf Basis der erhobenen Daten gingen die Autoren auf gedächtnisbezogene Kompetenzen ein und untersuchten ebenfalls die interindividuelle Stabilität. Für den Bereich der Entwicklung der Intelligenz wurde ein deutlicher Fähigkeitszuwachs vom Vorschulalter an bis in das späte Schulalter beschrieben. Die Entwicklungsverläufe ähneln sich. Aber im Vorschulalter beobachtete interindividuelle bestehende Kompetenzunterschiede bleiben jedoch annähernd erhalten. Die Autoren resümieren, dass interindividuelle Unterschiede im intellektuellen Niveau durch schulische Interventionen nur begrenzt beeinflussbar sind. Rost (2010) geht dagegen davon aus, dass die hohe interindividuelle Stabilität erst mit dem 3. Schuljahr festgestellt werden kann. Über die interindividuelle Stabilität kognitiver Fähigkeiten von Kindern des ersten und zweiten Schuljahres berichtete Otto (2012). Für den betrachteten Zeitraum konnte auf korrelativer Basis festgestellt werden, dass die Intelligenzleistungen (verbale, quantitative und nonverbale Subskalen sowie die als Gesamtwert ermittelte kognitive Fähigkeit) des ersten und zweiten Messzeitpunktes sehr signifikant miteinander korrelierten (Otto, 2012). Die Gesamtleistung korrelierte insgesamt am höchsten mit .77, gefolgt von der quantitativen Subskala der ersten und zweiten Klassenstufe mit .71 (Otto, 2012). Die ermittelte Positionsstabilität zeigte in keiner Skala eine signifikante Veränderung der Rangpositionen. Dennoch waren innerhalb der Gruppe Veränderungen der Rangplätze zu verzeichnen. Die auf manifester Ebene durchgeführten Regressionsanalysen zeigten, dass Intelligenzleistungen des ersten Messzeitpunktes die Intelligenzleistungen des zweiten Messzeitpunktes mit .47 vorherzusagen vermochten (Otto, 2012). Die quantitativen Fähigkeiten konnten die quantitativen Fähigkeiten des zweiten Messzeitpunktes mit .69 vorhersagen, während verbale und nonverbale Fähigkeiten die jeweiligen Fähigkeiten zum zweiten Messzeitpunkt mit .43 und .33 geringer zu prognostizieren vermochten (Otto, 2012).

In der Untersuchung zu den Effekten von Diagnoseförderklassen konnte von Hartke, Koch und Blumenthal (2010) hinsichtlich der Stabilität der Schulleistungen festgestellt werden, dass nicht alle zu Beginn ihrer Erhebung als schulisch gefährdet eingestuften Kinder eine Schulleistungsstörung entwickelten. Trotz der um ein Jahr längeren Beschulung in den DFK-Klassen erreichen mehr gefährdete Grundschüler (65 %) als DFK-Schüler (46 %) einen unauffälligen Befund. Die Rechenleistung der Kinder erreicht in der DFK-Beschulung erst nach drei Jahren das Niveau, das vergleichbare Kinder nach zwei Jahren in der

Grundschule besitzen. Hinsichtlich der Rechenleistung wirke sich demnach allein eine verlängerte Lernzeit nicht positiv aus. Gezielte Interventionen und Präventionsmaßnahmen erscheinen den Autoren dagegen geeigneter, so Hartke, Koch und Blumenthal (2010).

In der differenzierten Betrachtung stellten Schneider et al. (2014) anhand des LOGIK-Datensatzes fest, dass Kinder, die im Alter von 7 Jahren niedrige kognitive Fähigkeiten hatten, eine höhere interindividuelle Stabilität aufweisen und auch häufiger eine nichtakademische Berufslaufbahn wählen. Personen mit früh diagnostizierten hohen kognitiven Fähigkeiten streben dagegen später zumeist auch einen akademischen Abschluss an. Bemerkenswert jedoch sind zwei Ergebnisse, wie Schneider et al. (2014) betonen. Zum einen gab es einige Personen mit einem Intelligenzquotienten kleiner 85, die später einen Intelligenzwert von 120 aufwiesen, aber auch Personen mit hohen kognitiven Fähigkeiten, die später einen auffallend niedrigen Bildungsabschluss erzielten. Schneider et al. (2014) diskutieren in diesem Zusammenhang Umwelteinflüsse, die auf die interindividuelle Stabilität Einfluss nehmen. Dies jedoch bezweifeln Bornstein, et al. (2006), die anhand einer bemerkenswerten Datenbasis die kognitiven Fähigkeiten von Säuglingen erfassten und diese mit deren Leistungen im Alter von vier Jahren verglichen. Sie erfassten visuelle Aufmerksamkeit und verschiedene neurobiologische Merkmale und konnten feststellen, dass die geistige Entwicklung der Kinder weniger von Umweltbedingungen und familiären Faktoren beeinflusst ist als angenommen. Sie vermuten eher intrapersonelle Ursachen auf neuronaler Ebene. Vierjährige mit hohen kognitiven Fähigkeiten konnten bereits im Säuglingsalter leichter und realitätsnaher auf Erinnerungen zugreifen. Als Erklärung werden Einflüsse der Informationsverarbeitungsgeschwindigkeit und Myelinisierungsunterschiede angenommen. Ergebnisse hoher interindividueller Stabilität berichten auch Rönnlund et al. (2015), die interindividuelle kognitive Fähigkeiten im Sinne einer allgemeinen kognitiven Fähigkeit g im Alter zwischen achtzehn und fünfundsechzig Jahren bei 265 Männern untersuchten. Diese absolvierten in Fünfjahresabständen bis zum Alter von 65 Jahren einen bereichsspezifischen kognitiven Fähigkeitstest. Die Autoren fanden sehr hohe Langzeitstabilitäten und berichteten auf der Basis der Modellierung von Strukturgleichungsmodellen (SEM) Stabilitätskoeffizienten von .90 oder höher. Auch auf der latenten Ebene konnten bis zu 90 % der Varianzanteile 32 bis 47 Jahre später erklärt werden. Ferrer und McArdle (2004) widmeten sich der Stabilität kognitiver Fähigkeiten und schulischer Leistung von der Kindheit bis zum frühen Erwachsenenalter. Anhand linearer Modelle stellten sie fest, dass fluide Denkfähigkeiten ein bedeutender Prädiktor für Veränderungen mathematischer und allgemeiner schu-

lischer Leistungen sind. Allerdings wiesen sie darauf hin, dass sich die Beziehungen zwischen kognitiven Fähigkeiten und akademischen Leistungen komplexer gestalten als sie mit der Investitionshypothese von Cattell (1971) erklärt werden können. Denn Forschungsergebnisse zeigen teilweise Entwicklungen der Stabilität der Intelligenz, die häufig in konträre Richtungen weisen. So werden im Zuge der Stabilität sowohl der Matthäus-Effekt beschrieben, der ein Auseinanderdriften der kognitiven Fähigkeiten im Schulalter erklären könnte, aber auch genau gegenteilige Effekte, die als Kompensationseffekte bezeichnet werden. Diese Effekte zeigen sich, wenn Kinder mit niedrigeren kognitiven Fähigkeiten durch Effekte der Beschulung eine Positionsveränderung nach oben über die Zeit vollziehen. Diese Ergebnisse stehen im Gegensatz zu dem Matthäus-Effekt. So gingen beispielsweise Schroeders et al. (2016) der Frage nach, inwieweit Annahmen des Matthäus-Effekts bestätigt werden können oder aber ob durch schulische Bildung Kompensationseffekte bei Kindern mit geringeren kognitiven Fähigkeiten zu verzeichnen sind, wie sie von Hambrick (2004), Hambrick und Engle (2002) berichtet werden. Sie untersuchten 1102 Grundschüler der dritten Klassenstufen, die zu zwei Testzeitpunkten im Abstand von fünf Monaten einen Test der fluiden Denkfähigkeiten absolvierten. Zum Zusammenhang zwischen quantitativen KFT-Skalen und den mathematischen Schulleistungen in der ersten Klassenstufe (DEMAT 1) konnte anhand der Daten der Mecklenburger Längsschnittstudie eine Korrelation auf manifester Ebene in Höhe von .51 berichtet werden (Otto, 2012). Der höchste signifikante Zusammenhang zeigte sich zwischen der Gesamtleistung des KFT von Abou-Koura und Perleth (2005) im ersten Schuljahr zur Gesamtleistung im DEMAT 1 von Krajewski, Küspert und Schneider (2002) in Höhe von .52 (Otto, 2012). Für die zweite Klassenstufe konnte auf manifester Ebene eine Korrelation in Höhe von .63 zwischen der quantitativen Subskala des KFT von Abou-Koura und Perleth (2005) und der Leistung im DEMAT 2 von Krajewski, Liehm und Schneider (2004) ermittelt werden. Die Korrelation des kognitiven Gesamtwertes der zweiten Klassenstufe zur mathematischen Gesamtleistung im DEMAT 2 zeigt eine Höhe von .60. In der Betrachtung über zwei Messzeitpunkte bestand eine Korrelationshöhe von .56 zwischen der quantitativen Skala des KFT und der Mathematikleistung im DEMAT 2, während die kognitive Gesamtfähigkeit des ersten Jahres einen signifikanten Zusammenhang zur DEMAT-Leistung des zweiten Zeitpunktes in Höhe von .57 zeigte (Otto, 2012).

Ein bedeutendes Ergebnis der von Otto (2012) durchgeführten Studie ist, dass sich bei einem Vergleich verschiedener Personengruppen anhand ihrer kognitiven Merkmale ein zunehmender Vorteil für die kognitiv stärkeren Kinder nachweisen ließ (Otto, 2012). Damit kann der von Stanovich (1986) beschriebene Matthäus-Effekt bestätigt werden. Weinert und

Helmke (1997) bezeichneten diese zunehmenden Lernvorteile Vorteile beim Wissenserwerb als kumulatives Lernen. Schroeders et al. (2016) erhoben in dem Schuljahr 2012-2013 Daten von hochbegabten Kindern, die an dem Hector Children's Academy Program (HCAP) teilnahmen. Dieses Programm fördert begabte Kinder in den naturwissenschaftlichen Fächern Wissenschaft, Technik, Ingenieurwesen und Mathematik außerhalb des regulären Lehrplans. Um das gesamte Spektrum kognitiver Fähigkeiten angemessen zu repräsentieren, erhoben sie Daten von zwei regulären Grundschulklassen aus Baden-Württemberg. Schroeders et al. (2016) konnten starke Kompensationseffekte für domänenspezifische Bereiche feststellen. Die fluiden Intelligenzanteile und auch die kristallisierten Intelligenzanteile der Schüler mit niedrigerem kognitivem Ausgangsniveau näherten sich zum zweiten Zeitpunkt stärker an als es für die Schüler mit höherem Ausgangsniveau der Fall war. Allerdings erhoben die Autoren die fluiden Fähigkeiten mit verbalen, numerischen und figuralen Items. Ob diese Items wirklich fluide Anteile erfassen, wie von den Autoren berichtet, oder nicht eher, wie von den Testautoren der kognitiven Testreihe berichtet, eher schulnahe kristallisierte Anteile – mit Ausnahme der nonverbalen Skala – (Perleth, 1997) erfassen, bleibt eine offene Frage. Den Kompensationseffekt erklären Schroeders et al. (2016) mit Effekten des schulischen Kontextes. Kompensationseffekte steuern den personengebundenen kumulativen Effekten entgegen, in dem schulische Grundlagen für alle Schüler erreicht werden können, welche dann nicht weiter ausgebaut werden, sondern im Niveau stabil bleiben. Dadurch können die langsamer Lernenden das gleiche Niveau erreichen, wenn auch später (Bast & Reitsma, 1998). In dieser Zeit lernen die fähigeren Personen, in anregendem Unterricht, andere Inhalte bzw. spezialisieren sich, oder aber sie schöpfen im weniger anregenden Unterricht ihr kognitives Potential nicht aus. So könnte der Kompensationseffekt ein Resultat von Deckeneffekten sein, der den Schülern keine Gelegenheit gibt, ihr tatsächliches Leistungsvermögen zu zeigen (Schroeders et al., 2016).

2.2.3.2 Autoregressive Effekte mathematischer Schulleistungen über die Zeit

In Bezug auf die Stabilität mathematischer Schulleistungen ergibt sich folgendes Bild.

Schneider, Bullock und Sodian (1998) zeigen eine hohe Stabilität der Schulleistungen, gemessen allerdings mittels Schulnoten (Mathematik und Deutsch) und auf manifester korrelativer Basis. Aufgrund der berichteten Korrelationen zwischen kognitiven Fähigkeiten und Schulnoten gehen sie aber davon aus, dass die psychometrische Intelligenz als bedeutender

Prädiktor schulischer Leistungen fungiere. Sie gehen also davon aus, dass die Stabilität aufgrund kognitiver Fähigkeiten zustande kommt und stellen nicht die Frage, inwieweit gegenseitige kreuzverzögerte Effekte zu den Ergebnissen führten.

Auch für die relative Stabilität von mathematischen Schulleistungen gibt es Hinweise, wie sie beispielsweise von Krajewski (2003) und Stern (2008) sowie Weinert und Helmke (1997) berichtet werden. Als Ergebnis der LOGIK-Studie konnte gezeigt werden, dass Mathematikleistungen stabil zu sein scheinen (Stern, 2008). Kinder mit Schwierigkeiten im Lösen von Testaufgaben in der zweiten Klasse hatten auch später bei Algebra-Aufgaben Probleme.

In der differenzierten Betrachtung wird diskutiert, ob die Stabilität von Mathematikleistungen für bestimmte Personengruppen unterschiedlich hoch ist. So untersuchten Martin et al. (2012) 144 Grundschulkinder aus Tennessee und Texas mit mathematischen Schwierigkeiten und bildeten anhand der Daten Gruppen hinsichtlich der Schwere der Beeinträchtigung. Es gab keinen Unterschied, ob die Kinder nur leichte oder gravierende mathematische Lernschwierigkeiten hatten. Die Stabilität über zwei Jahre blieb hoch. Auch Stern verweist darauf, dass mathematische Kompetenzen als Ergebnis einer stetigen Auseinandersetzung mit intellektuell anregenden Aufgaben verbessert werden können (Stern, 2008).

Silver et al. (1999) dagegen warnten davor, die Einschätzung von Teilleistungsschwierigkeiten aufgrund einer einmaligen Messung als dauerhafte Teilleistungsergebnisse zu interpretieren, denn er konnte anhand der Daten von 80 Kindern zeigen, dass sich die Bereiche, in denen Kinder Schwierigkeiten haben, verändern, nicht jedoch der Fakt, dass dauerhafte Verbesserungen selten sind.

Die Stabilität von Mathematikleistungen gilt bereits im Kindesalter als hoch. Grube und Barth (2004) und auch Helmke und Schrader (2010) gehen davon aus, dass kognitive Fähigkeiten einen großen Anteil an der Qualität von Mathematikleistungen besitzen. Dabei werden diese vom Arbeitsgedächtnis, metakognitiven Komponenten, dem bereits in der Vorschulzeit erworbenen Vorwissen sowie dem Niveau des Zahlen- und Mengenverständnisses determiniert (Grube, 2006a). Die Beziehungen zum Arbeitsgedächtnis werden recht unterschiedlich berichtet. Weberschock und Grube (2006) weisen auf eine mathematikaufgabenspezifische, unterschiedliche Beteiligung der Arbeitsgedächtniskomponenten hin. Befunde von Hasselhorn und Grube (2006) bestätigen dies.

Doch obwohl die Bedeutung des Unterrichts und dessen Moderatoren für die Schülerleistung unbestritten ist, ist dennoch nicht geklärt, wie die hohe zeitliche Stabilität großer

interindividueller Leistungsunterschiede für das Fach Mathematik erklärt werden kann (Stern, 1997, in Weinert & Helmke, 1997). Zu allen Messzeitpunkten der Untersuchung von Stern zeigt sich eine Standardabweichung von 30 % bis 40 %. Die Höhe der benachbarten Korrelationen liegt bei .68 bis .79. Sogar über mehrere Messzeitpunkte hinweg zeigt sich eine Korrelation von .53. Sie bedeutet, dass leistungsschwache Kinder bei Eintritt in die Schule auch zum Ende der Grundschulzeit überdauernd zu den leistungsschwächeren Kindern gehören.

Wie im Kapitel Stabilität dargelegt wurde, existiert eine hohe zeitliche Stabilität großer interindividueller Leistungsunterschiede für das Fach Mathematik, die bedeutet, dass Kinder, die beim Eintritt in die Schule leistungsschwach sind, überdauernd auch zum Ende der Grundschulzeit zu den leistungsschwächeren Kindern gehören. Im Rahmen dieser Ergebnisse wurden beispielsweise instruktionale Unterschiede diskutiert, so etwa, ob die Vorstellungen der Lehrer über den Erwerb mathematischer Kompetenzen einen Einfluss auf die Leistungsfortschritte von Kindern haben. Hier zeigte sich ein positiver Effekt einer konstruktivistischen Vorstellung gegenüber der Vorstellung, der Erwerb mathematischer Kompetenzen folge vornehmlich einem rezeptiven Lernvorgang (Stern, 1997). Lehrer mit einer konstruktivistischen Grundhaltung über den Erwerb mathematischer Kompetenzen besitzen demnach eine höhere Sensibilität für noch unsicher verfügbares Wissen (Stern, 1997).

Hinsichtlich der unerwünscht hohen zeitlichen Stabilität der interindividuellen Unterschiede in den Mathematikleistungen wurde auch der Einfluss der Schulart erwogen. Ditton und Kreckler (1995) gingen dieser Frage nach und konnten zeigen, dass der Einfluss der Schulart auf die Schülerleistungen bei etwa 2,5 % liegt. Andere Studien, wie beispielsweise von Jencks et al. (1972), führt nur 1 % der Leistungsvarianz auf Schuleffekte zurück.

2.2.3.3 Autoregressive Effekte – Zusammenfassung, Ableitung und Hypothesen

In den oben aufgeführten Berichten zu vorliegenden Studien ist deutlich geworden, dass es keine eindeutigen Ergebnisse zur interindividuellen Stabilität kognitiver Fähigkeiten über das Grundschulalter gibt. Im Rahmen der Sichtung vorliegender Studien zu diesen Fragen fiel auf, dass nur wenige längsschnittliche Forschungsergebnisse zur Stabilität von kognitiven Fähigkeiten für das Grundschulalter existieren. So liegen bisherigen Untersuchungen für diesen Zeitraum nur maximal zwei Messzeitpunkte zugrunde, wodurch die Aussagekraft eingeschränkt wird, denn die Häufigkeit der Messungen spielt eine zentrale Rolle. Einmalige Messungen erlauben zwar auf korrelativer Ebene einen Blick auf die Zusammenhänge

zwischen kognitiven Fähigkeiten und mathematischen Schulleistungen, jedoch lassen sie keine Rückschlüsse von Einflüssen früherer Zeitpunkte auf spätere Zeitpunkte zu.

Häufig basieren die Untersuchungen auch auf kleinen Stichprobengrundlagen, oder aber umfassen wesentlich längere Zeiträume. Bei Untersuchungen über große Zeitspannen hinweg ist auch der Einsatz verschiedener Testverfahren problematisch, da Messfehlereinflüsse die Ergebnisse beeinträchtigen. Auch wenn hohe Stabilitätswerte über die Lebenszeit berichtet werden, ist es möglich, dass die Ergebnisse von der Art der statistischen Spezifikation abhängen, ob also beispielsweise Daten auf manifester oder auf latenter Ebene analysiert wurden. Weiterhin ist für die Aussagequalität der Ergebnisse das zugrundeliegende theoretische Modell der Struktur der Intelligenz entscheidend, denn es macht einen Unterschied, ob eine multifaktorielle Annahme besteht oder ein Testverfahren auf der Annahme eines allgemeinen Intelligenzfaktors g beruht. Zudem ist die Qualität der Testinstrumente ausschlaggebend. Vor dem Hintergrund dieser Einschränkungen sind für die Betrachtung der Stabilität innerhalb der Kindheitsphase die Stabilitätsbefunde insgesamt uneinheitlich.

So stehen Positionen, die von kumulativem Lernen und einem Matthäus-Effekt ausgehen, Positionen gegenüber, die von einem Kompensationseffekt überzeugt sind. Für präzisere Aussagen sind zum einen mehr als zwei Messzeitpunkte notwendig, als auch eine hinreichend große Stichprobe sowie ein methodisch fundiertes Analysedesign. Bislang wurden Aussagen über die Stabilität über vier Messzeitpunkte im Grundschulalter noch nicht berichtet. Die zugrundeliegende Datenstruktur von vier Messzeitpunkten im jährlichen Abstand ermöglicht eine Betrachtung der *interindividuellen* Stabilität der untersuchten Konstrukte im Grundschulalter. So sind die eingesetzten Modelle (siehe Kapitel Analysestrategien) geeignet, Aussagen über die Rangreihenstabilität zu erlauben, jedoch können sie nichts über intraindividuelle Mittelwertverläufe und über das mittlere Wachstum in der Stichprobe (Niveauveränderungen) aussagen. Die vorliegende Arbeit versucht daher zu beantworten, inwieweit die erfassten Konstrukte kognitive Fähigkeiten und mathematische Leistungen im Grundschulalter als stabil zu betrachten sind. Die Beantwortung besitzt zum einen eine zentrale Bedeutung für die pädagogische Praxis und zum anderen auch für die Grundlagenforschung zur Stabilität von Leistungsmerkmalen. Beispielsweise könnte beantwortet werden, ob Leistungen der ersten Klassenstufe noch Leistungen gegen Ende des Grundschulalters vorhersagen können.

In der 2. Fragestellung steht die also die interindividuelle Stabilität der erfassten Konstrukte über die Zeit im Mittelpunkt. Es soll beantwortet werden, ob sich die interindividuelle Stabilität kognitiver Leistungen und mathematischer Schulleistungen anhand der

längsschnittlichen Untersuchung von vier Messzeitpunkten in der Grundschulzeit von Kindern der 1. bis 4. Klassenstufe tatsächlich nachweisen lässt. Fraglich ist, inwieweit sich das Konstrukt der Intelligenz unter Berücksichtigung des Einflusses von Messfehlern gegen Ende des Grundschulalters zuverlässig aus früheren Messungen vorhersagen lässt und inwiefern sich mathematische Schulleistungen aus früher gemessenen mathematischen Schulleistungen ableiten lassen. Auch ob eine präzise Aussage zur Stabilität beider Konstrukte in der separaten Betrachtung erlaubt sind, ist zu fragen. Bislang sind keine Untersuchungen bekannt, die sich dieser Frage über einen Zeitraum von vier Jahren widmen. Auch zu der Stabilität der quantitativen Fähigkeiten existieren auf latenter Ebene über den betrachteten Zeitraum noch keine Ergebnisse. Auch der Erkenntnisgewinn in der Debatte um die Nachweisbarkeit des Matthäus-Effekts bzw. eines Kompensationseffekts kann nicht unterschätzt werden.

Die dritte Fragestellung, die nach der Stabilität der erfassten Konstrukte kognitive Fähigkeiten und mathematische Schulleistungen über einen Vierjahreszeitraum im Grundschulalter fragt, hilft, präzise Antworten zur Stabilität der Konstrukte kognitive Fähigkeiten und mathematische Schulleistungen zu finden. Sie leistet damit einen Beitrag zu noch unbeantworteten Fragen der aktuellen Intelligenz- und Schulleistungsforschung.

Zur Feststellung der interindividuellen Stabilität kognitiver Leistungen und mathematischer Leistungen werden mehrere Messungen aus zeitlichen Abfolgen im Längsschnitt verglichen.

Hypothese 3a:

Es wird erwartet, dass die Stabilität kognitiver Fähigkeiten über den betrachteten Zeitraum hoch ist.

Hypothese 3b:

Es wird erwartet, dass die Stabilität mathematischer Schulleistungen über den betrachteten Zeitraum hoch ist.

2.2.4 Fragestellung 4 – Kreuzverzögerte Effekte mathematischer Schulleistungen und kognitiver Fähigkeiten

Der Zusammenhang zwischen Intelligenz und akademischem und schulischem Erfolg ist theoretisch gut etabliert (Naglieri & Bornstein, 2003). Allerdings ist der bisherigen Forschungslage nicht eindeutig zu entnehmen, ob schulische Leistungen kognitive Fähigkeiten beeinflussen, kognitive Fähigkeiten schulische Leistungen bestimmen oder ob und inwiefern die Effekte gegebenenfalls eine gegenseitige Wirkung entfalten. Im Kontext der Schulleistungsforschung wird zumeist angenommen, dass entsprechendes Wissen und mathematische Fähigkeiten vornehmlich in der Schule erworben werden (Geary, 1993; Köller & Baumert, 2002) und damit die Basis des schulischen Erfolgs bilden. Für mathematische Schulleistungen wird seitens der pädagogisch-psychologischen Forschung ergänzt, dass sich für das Fach Mathematik die kognitiven Fähigkeiten als dominierender Prädiktor von Mathematikleistungen (Weber et al., 2015) erweisen. Auch Hinweise, inwiefern vielmehr in der Schule erworbenes Wissen wiederum kognitive Fähigkeiten bestimmt, werden diskutiert. Schulerfolg wird jedoch auch von anderen Bedingungen beeinflusst. Gesellschaftliche (Bronfenbrenner, 1986, Esser, 1999a) institutionelle (Berger & Kahlert, 2013), familiäre Faktoren (Perleth, 1997; Helmke & Weinert, 1997) sowie nicht-kognitive persönliche Moderatoren (Heller, 1991, Perleth, 1997; Spinath, 2006) werden als Determinanten angeführt.

In dem folgenden Kapitel soll nun der Versuch unternommen werden, diese entscheidende Frage zu klären. Zuerst werden kurz die einflussnehmenden moderierenden Faktoren angesprochen, die in der Literatur als nicht zu vernachlässigend gelten. In nächsten Schritt werden dann die Theorie-Positionen der Forscher vorgestellt, die kognitive Fähigkeiten vornehmlich durch die Lernumgebung Schule verortet sehen. Hier gilt die besondere Aufmerksamkeit dem besonders wichtigen schulisch erworbenen Vorwissen. Im Anschluss steht die Position der Autoren im Vordergrund, die von einem prädominanten Einfluss der kognitiven Fähigkeiten auf die Schulleistungen ausgehen. In diesem Zusammenhang wird auf die Bedeutung des Vorwissens im Sinne der Expertiseforschung eingegangen. Da das Vorwissen eine grundlegende Bedeutung sowohl für die in der Schule erworbenen mathematischen Leistungen als auch die kognitiven Fähigkeiten hat, wird im Anschluss darüber diskutiert, ob es eine gegenseitige Beziehung zwischen den Konstrukten gibt. In diesem Zusammenhang werden neue Ansätze vorgestellt, die von einer gegenseitigen Beeinflussung von kognitiven Fähigkeiten und mathematischen Schulleistungen ausgehen. Da es erst

wenige Forscher wie zum Beispiel Perleth gibt, die sich mit dieser Annahme auseinandergesetzt haben, sind die vorliegenden Theorien im Rahmen dieser Arbeit umso wichtiger. Die anschließende kurze Zusammenfassung führt zur Ableitung der vierten Fragestellung der vorliegenden Arbeit.

Zunächst aber werden die oben bereits angesprochenen, die Konstrukte moderierenden Faktoren der Vollständigkeit halber eingangs kurz beleuchtet, denn es besteht eine zumindest grundsätzliche Einigkeit, dass eine multifaktorielle Beeinflussung der Konstrukte nicht ausgeschlossen werden kann. Zu ihnen gehören sowohl personeninterne kognitive (z. B. Arbeitsgedächtnis, Verarbeitungsgeschwindigkeit) und nicht-kognitive (z. B. Motivation, Kausalattribution) Persönlichkeitsmerkmale als auch Umweltmoderatoren. Unter ihnen identifizieren Kaufman, DeYoung, Gray, Brown und Mackintosh (2009) das Arbeitsgedächtnis, assoziatives Lernen und Verarbeitungsgeschwindigkeit als die wichtigsten kognitiven Prozesse, die Einfluss auf akademische Leistungen haben. Sie beschäftigten sich mit den kognitiven Fähigkeiten *g* und den Leistungsmerkmalen unter dem Aspekt, inwieweit diese Konstrukte hoch ähnlich oder sogar identisch seien. Ihre Literaturrecherche zeigt, dass vielfach bisher berichtete Studien ein Konglomerat verschiedener Intelligenztestbatterien analysieren. Allein aufgrund dieser Tatsache entstehen Ungenauigkeiten, da die Tests beispielsweise unterschiedliche hierarchische Organisationen der einzelnen kognitiven Fähigkeiten annehmen, aber in der Gesamtanalyse oftmals vermischt werden. So werden letztlich ohnehin allgemeine kognitive Fähigkeiten gemessen. Und obgleich die bestehenden Korrelationsbeziehungen schulischer Leistungsmaße und kognitiver Fähigkeitstests deutlich sind, betont Kaufmann (2012) dagegen noch ein bislang wenig beachtetes Problem der Beziehung zwischen kognitiven Fähigkeiten (COG-*g*) und akademischen Leistungsmerkmalen (ACH-*g*) (Kaufmann, 2102). Er führt an, dass mit zunehmendem Alter nicht-kognitive Merkmale wie Motivation und Ausdauer (Jensen, 1998; Perleth, 1997) eine wesentliche Rolle spielen. Kaufman et al. (2012) beschließen ihre Untersuchungen der Beziehungen zwischen kognitiven Fähigkeiten und Schulleistungsmaßen mit der Aussage, dass keine Identität zwischen beiden vorliegt. Obwohl sie ähnlich hohe Korrelationskoeffizienten wie Deary et al. (2007) mit .81 von .77 bis .88 nachwiesen, fanden sie auch starke Belege für den Einfluss anderer Konstrukte (Deary, 2007). Während die Korrelationen zeigen, dass etwa 50 % bis 60 % der Varianz der Schulleistungen durch kognitive Faktoren *g*-Faktor statistisch erklärt werden können, entfiel ein ebenso hoher Anteil der Varianz nicht auf eine allgemeine kognitive Fähigkeit *g*. Ein Teil dieser verbleibenden Varianz kann durch Mess-

fehler erklärt werden, aber Deary stellt fest, dass die „Nicht-g-Faktoren“ wesentliche Auswirkungen auf den Bildungserfolg besitzen. Hierbei handele es sich um Engagement, Schülerpersönlichkeit, Motivation, Anstrengung und das Ausmaß der elterlichen Unterstützung sowie die Bereitstellung geeigneter Lernerfahrungen, die Qualität der Lehre, das Schulklima und andere mögliche Faktoren. Die Studie von Vock, Preckel und Holling (2011) bewies den wesentlichen Einfluss mentaler Geschwindigkeit und des Kurzzeitgedächtnisses als Fähigkeit zu komplexen kognitiven Leistungen wie Argumentieren und divergentes Denken auf Leistungsmaße. Jensen (1998) stellte fest, dass der erfolgreiche Abruf früherer Wisseninformationen zum Einsatz aktueller Anforderungen von Prozessen der Informationsverarbeitung abhängig ist. Luo, Thompson und Detterman (2006) bestätigten die dominante Rolle mentaler Geschwindigkeit in der Beziehung zwischen kognitiven Fähigkeiten und akademischer Leistung, aber die kristallisierten Intelligenzanteile (Bildungswissen) sind als Prädiktoren schulischer Leistung bestätigt worden (Luo et al., 2006). Sie sprachen sich für eine Moderatorrolle bzw. für das Verständnis als potentiell identische Fähigkeiten aus (in Kaufmann, 2012). Grundlegende Prozesse der Informationsverarbeitung und kristallisierte Intelligenz zusammen können etwa 60 % der schulischen Leistung erklären. Die Autoren stellen damit den Wert der Intelligenztests, die lediglich fluide Fähigkeiten messen können, infrage und empfehlen den Einsatz von Testverfahren, die methodenspezifische und prozessspezifische Fähigkeiten sowie kristallisierte Intelligenzanteile messen können (Luo, et al., 2006). Luo et al. (2006) konzentrieren sich auf die Verarbeitungsgeschwindigkeit und die Arbeitsspeicheraufgaben als Prädiktoren schulischer Leistungen und stellen eine erhebliche Erklärungskraft fest. Sie bestätigen den weitreichenden Einfluss kristallisierter Intelligenz als akkumuliertes Wissen und damit deren Eignung für die Vorhersage schulischer Leistung. Grundlegende Prozesse und kristallisierte Intelligenz machen zusammen 60 % des Erklärungsanteils aus. Heller (1991, 1997); Weinert und Helmke (1997) oder auch Helmke, Rindermann und Schrader (2008) fragten in weiteren Studien nach den Einflüssen zwischen kognitiven Fähigkeiten und Schulleistungen und stellten Analyseergebnisse, die auf korrelativer Ebene den Gymnasialerfolg von ehemaligen Grundschulkindern untersuchten, in das Zentrum ihrer Forschungsbemühungen. Sie konstatierten, dass neben kognitiven Leistungen andere Einflussfaktoren unmittelbar die Schulleistungen beeinflussen. Spinath et al. (2006) untersuchten den Einfluss von motivationalen Aspekten auf die Leistung von neunjährigen Grundschulkindern in Großbritannien. Dabei absolvierten 1678 Kinder einen allgemeinen Intelligenztest, während die Lehrer die Leistungsbewertungen in Form von No-

ten zur Verfügung stellten. Sie ermittelten eine mittlere Korrelation zwischen den allgemeinen kognitiven Fähigkeiten und der Schulleistung in Höhe von .50 und konnten damit einen moderierenden Effekt der Motivation nachweisen. Auch Winther (2006) und Steinmayr und Spinath (2009) untersuchten die Motivationsabhängigkeit in Lernprozessen. Die Bedeutung der Verfügbarkeit und Qualität der formalen (z. B. in der Schule) und informellen (z. B. in Familie oder Gemeinschaft) Lernerfahrungen wie auch von persönlichen Faktoren wie Interessen, Motivation und Persistenz wird ebenfalls von Kvist und Gustafsson (2008) berichtet. Auch Lynn und Vanhanen (2012), die den Zusammenhang zwischen kognitiven Fähigkeiten und Schulleistungen anhand hoher Korrelationen in Höhe von .50 bis .70 belegten, geben zu bedenken, dass die Ergebnisse auch durch andere Faktoren moderiert werden. Sie berichten, dass der Aufklärungsanteil anderer Faktoren zwischen 3 und 19 Prozent lag (Lynn und Vanhanen, 2012). Lynn und Vanhanen (2012) schätzten die Zusammenhänge zwischen dem Niveau der Intelligenz in 108 verschiedenen Ländern mit den Bildungsabschlüssen von Schülern in Mathematik, Naturwissenschaften und Leseverständnis. Die Korrelation zwischen dem Bildungsstand eines Landes und den Intelligenzscores seiner Einwohner lag um .92, doch eine teilweise geringe Datenqualität schränkt die Aussagekraft der Studienergebnisse ein. Auch der intensiven Beschäftigung mit Lerninhalten kommt im Zusammenhang mit schulischen Leistungen große Aufmerksamkeit zu. So gehen zahlreiche Interventionstrainings zum Ausgleich von bereichsspezifischen Lern- bzw. Leistungsschwächen davon aus, dass durch erfolgreiche intensive Auseinandersetzung mit dem Thema ein Anstieg der Leistungen zu erzielen ist (Lorenz, 2009a). Lerntrainings werden auf diesem konzeptionellen Ansatz zur Unterstützung bereitgestellt, wie beispielsweise das Mathematikförderprogramm für Kinder „Elfe und Mattis“ von Lehnhard und Lehnhard (2009). Forscher thematisierten deshalb die Frage der Trainierbarkeit. Im Rahmen der Forschungsfrage, inwieweit Kontrollprozesse des Gedächtnisses und Zentrierung von Problemlösungen ursächlich für schulische Leistungsunterschiede sind. Es nehmen auch kognitive Merkmale wesentlichen Einfluss auf schulische Leistungen. So berichtet Heller (1991) von motivationalen Faktoren, der Einfluss des Lernverhaltens wird von Schäfer und McDermott (1999) dargelegt, und Hattie hebt die Qualität der Lehrereinflusses hervor (2012). Weiter weisen Berichte der Unterrichtsforschung auf den Prädiktor "Klassenführung" hinsichtlich schulischer Leistungsentwicklung hin (Blum & Krauss, 2007), der über entsprechende Erweiterung des fachdidaktischen Lehrerwissens zu verändern sei (Hattie, 2012). Insbesondere Hattie (2012) beispielsweise thematisiert eindrucksvoll die Wirkung

des Lehrerhandelns im Unterricht und gibt umfangreiche Empfehlungen, wie dieses optimiert werden kann, da Lehrerhandeln eine leistungsförderliche Wirkung auf Schülerkompetenzen habe. Diese Metaebene betrachteten Baumert et al. (2009a, 2009b) eingehender, indem sie das Professionswissen von Lehrkräften sowie den Einfluss kognitiv aktivierenden Mathematikunterrichts auf die Entwicklung mathematischer Kompetenz untersuchten. Auch in dem von Klieme, Schümer und Knoll (2001) modellierten „Basismodell guten Unterrichts“, das die empirischen Befunde der TIMSS-Studie aufnimmt, stehen Leistung und konzeptionelles Verständnis in Abhängigkeit der Qualität des Unterrichts, die über kognitive Aktivierung, Klassenführung und unterstützendes Unterrichtsklima definiert werde. Herrnstein und Murray (1994) weisen außerdem auf den Einfluss der Klassenstruktur auf die kognitiven Fähigkeiten hin (in Brinch & Galloway, 2011). Die Beeinflussbarkeit durch Bildungsinvestitionen in der frühen Kindheit wird beispielsweise von Neisser et al. (1996) thematisiert. In zahlreichen (Schul-)Leistungsmodellen wird der Einfluss weiterer Faktoren auf die Schulleistungen thematisiert (Ceci, 1991; Perleth, 1997). Die Instruktionsqualität des Unterrichts, die Klassenstärke, und weitere, im Allgemeinen auf der schulinternen Ebene interagierende Faktoren bestimmen dabei unterstützend die resultierende individuelle Schulleistung. Ein relativ hoher Konsens erkennt die Bedeutung von Umweltfaktoren in der frühen Kindheit an, wie wirtschaftliche, sozioökonomische Bedingungen oder auch Interventionen, die von der Familie oder frühen Bildungseinrichtungen initiiert werden. Dem Anregungsgehalt der familiären und schulischen Lernumwelt sowie den nicht kognitiven Persönlichkeitsmerkmalen (z. B. Perleth, 1997) wird also große Bedeutung beigemessen. Neben Perleth (1997) und Ceci (1991), die die gesellschaftlichen und familiären Einflüsse thematisierten, berichten Akukwe und Schroeders (2016) neueste Ergebnisse. Sie untersuchten sozioökonomische, kulturelle, soziale und kognitive Faktoren in Bezug auf die Biologiekompetenz der Klassenstufe neun in Deutschland. Dazu wurden 543 Eltern zu ihren sozioökonomischen, kulturellen und sozialen Hintergründen befragt. Die Ergebnisse ihrer Analysen ergaben eine (geringe), jedoch vorhandene Vorhersagemöglichkeit für die biologische Kompetenz der Kinder aufgrund der Anzahl der Bücher im Besitz der Familie und der biologischen Kompetenz der Eltern. Gustafsson und Balke (1993) versuchten die Zusammenhänge zwischen kognitiven Fähigkeiten und Schulleistungen auch auf latenter Ebene zu analysieren. Sie untersuchten die Beziehung über 17 verschiedene Fächer und stellten fest, dass eine allgemeine kognitive Fähigkeit g und g_c eine wesentliche Menge (40 %) der erläuterten Varianz der allgemeinen Leistungsfähigkeit ist. Auch hier wurden jedoch

keine validen Schulleistungstests eingesetzt, sondern Erhebungen auf der Basis von Schulnoten zugrunde gelegt. Das Ergebnis jedoch spreche für einen erheblichen Anteil nicht aufgeklärter Varianz. Diese wird erklärt durch bereichsspezifische Leistungsfaktoren. Resümierend lässt sich also sagen, dass der Einfluss vieler anderer Faktoren so beispielsweise die Schulform oder auch domänenspezifische Fähigkeiten (Gustafsson & Balke, 1993) sowie emotionale Faktoren, Verhaltensausrprägungen, Motivation und metakognitive Elemente und Klassenraumfaktoren, das Curriculum und die Instruktionsqualität, sowie familiäre Einflüsse und sozioökonomische Faktoren die Beziehung zwischen kognitiven Fähigkeiten und akademischer Leistung beeinflussen (Wang, Haertel & Walberg, 1993).

Wendet man sich nun der Frage zu, inwieweit schulische Bildung auf kognitive Fähigkeiten wirkt, gehen die Meinungen auseinander und werden kontrovers diskutiert. Einige Forscher vertreten die Ansicht, dass vornehmlich frühere Schulleistungen die späteren Schulleistungen bestimmen. Das individuelle Niveau von schulischen Leistungen wird zudem als relativ stabil betrachtet. Es wird also davon ausgegangen, dass gezeigte schulische Leistungen und das erworbene Wissen auf ähnlichem Niveau auch zu späteren Zeitpunkten erbracht werden können. Diese relative Konstanz von Schulleistungen über die Schulbesuchsdauer ist empirisch gut belegt (Stern, 1997). Dahinter steht implizit oder explizit die Annahme, dass entsprechende Fähigkeiten primär in der Schule erworben werden (Geary, 1995; Köller & Baumert, 2002) und sich in schulischen Leistungen zeigen. Zugleich wird angenommen, dass kognitive Fähigkeiten dabei eine eher weniger bedeutende Rolle spielen.

Eine andere Interpretation wird aber auch vertreten. Dem schulisch erworbenen Vorwissen, das schon ab dem Kindergartenalter vermittelt werden kann, wird dabei eine große Bedeutung zugewiesen. Theoretischer Hintergrund ist die Annahme, dass die schulische Bildung selbst ein Training der Denkfähigkeiten ist, eingebettet in eine günstige Lernumgebung, quasi ein Enrichmentangebot, wie Begabungsforscher die für die kognitive Fähigkeitsentwicklung förderlichen Angebote bezeichnen. Studienergebnisse zeigen, dass der Schulbesuch und seine Dauer und Intensität selbst Einfluss auf die Entwicklung der kognitiven Fähigkeiten von Kindern hat. So berichtet Geary (2000), dass sich primitives quantitatives Verständnis zwar auch ohne formale Schulbildung im frühen Kindesalter entwickelt, doch sei dieses nicht konstant und das Erlernen sekundärer quantitativer Fähigkeiten erfordere hoch organisierte, fokussierte Lehrmethoden (Geary, 2000). Geary betont die Beeinflussbarkeit späterer Bildungsleistungen durch Bildungsinvestitionen in der frühen Kindheit. Jensen (1998) ging der Frage nach, ob nicht schon vorschulisch erworbene Bildung

einen Einfluss auf die kognitiven Fähigkeiten von Kindern hat. Seine Ergebnisse offenbarten, dass Kinder, die eine vorschulische Bildung genossen hatten, ein vergleichsweise höheres kognitives Niveau besaßen. Jensen (2000) schlussfolgert daher, dass aus theoretischer Sicht hinsichtlich des Intelligenzkonstrukts erwartet werden kann, dass vorherige schulische Leistung die späteren Schülerleistungen beeinflussen, weil schulisches Lernen selbst ein anspruchsvolles kognitives Lernen sei (Jensen, 2000). Jensen (1998) argumentiert allerdings infolge der unpräzisen Erfassung sowie der erheblichen Anteile nicht aufgeklärter Varianz, dass die starke Korrelation zwischen intellektuellen Fähigkeiten und schulischer Leistung nicht bedeutete, dass man das eine oder andere Konstrukt voraussagen könne, da es andere Faktoren für die unerklärte Varianz zu berücksichtigen gebe. Eine abschließende Klärung der Ursache- und Wirkungsbeziehung bleibt schwierig, da diese Faktoren untereinander in Beziehung stehen und diese methodisch nur schwierig aufzuklären ist. Für die Analyse des Einflusses von mathematischem Vorwissen auf die kognitiven Fähigkeiten konnte eine Studie von Shanley (2016) recherchiert werden, die die Beziehung zwischen der mathematischen Leistung zum Zeitpunkt des früheren oder späteren Kindergartenbesuchs und den mathematischen Wachstumsraten in den späteren Grund- und Mittelschulen untersuchte. Shanley konnte bemerkenswerterweise keine statistisch signifikanten Beziehungen feststellen. Er nahm diese Untersuchungen anhand von Wachstumskurvenberechnungen vor und resümiert, dass vor allem eine präzise Messung und Modellierung von schulischen Leistungen für die pädagogische Praxis und auch für bildungspolitische Entscheidungen entscheidend sei. Es wird dennoch davon ausgegangen, dass entsprechende mathematische Fähigkeiten primär in der Schule erworben werden. Insofern ist zu schließen, dass auch und insbesondere die Schulbesuchsdauer und die Intensität eine große Rolle spielen. So konnte Mackintosh (2011) nachweisen, dass die Jahre der formalen Bildung einen direkten Zusammenhang mit den kognitiven Fähigkeiten aufweisen; sowie mit dem Einkommen und dem späteren Berufsstatus korrelieren. Die durchschnittlichen Korrelationen zwischen Intelligenz und diesen Kriterien hatten eine Höhe von .55, .60, .40, und .55. Diese Werte könnten erklären, dass Menschen mit höheren kognitiven Fähigkeiten ein höheres Bildungsniveau erreichen und damit auch einen höheren beruflichen Status mit höherem Einkommen erzielen. Einen weiteren Beleg für den Einfluss der Schulbildung auf die Intelligenz sehen auch Cahan und Cohen (1989). Sie berichten von beträchtlichen Wirkungen der Schulbildung auf kognitive Fähigkeiten. Sie zeigen, dass Schulbildung eine etwa doppelt so starke Wirkung habe, wie das chronologische Alter auf das Niveau kognitiver Leistungen wirke. Die Untersuchung von Härnqvist bestätigt den Befund, als dieser Jugendliche untersuchte,

die zunächst einen vergleichbaren Intelligenzwert besaßen, aber die Schule früher (im Alter von 13 Jahren) als ihre Kontrollgruppe verließen. Für jedes Jahr, das ihnen zum High School Abschluss fehlte, wurde ein Verlust von 1,8 Intelligenzpunktwerten verzeichnet (bis zu einer maximalen Differenz von fast 8 Intelligenzpunktwerten (1968). Gustafsson seinerseits untersuchte die Wirkung von schulischer Bildung auf kognitive Fähigkeiten und stellte auch die dominierende Wirkung von schulischer Bildung auf die kognitiven Fähigkeiten in das Zentrum der Untersuchungen (Gustafsson, 2001). Dabei stellte er auch die Frage in den Mittelpunkt, wie stark die potentiell vorhandenen Wirkungen sind. Weiter sollte offengelegt werden, welche Bereiche der Intelligenz beeinflusst werden. Gustafssons Studie basiert auf den Überlegungen der Längsschnittstudie von Härnqvist (1968a, 1968b), in der 10 % der schwedischen Bevölkerung im Alter von 13 Jahren eine Testbatterie zur Messung kognitiver Fähigkeiten absolvierten. Eine zweite Erhebung erfolgte in der Klassenstufe 7 bis 9 und eine dritte in der Klassenstufe 10 bis 12.

Personen, die eine ein Jahr längere Schulbildung hatten, schnitten in Intelligenztests um durchschnittlich zwei Punkte besser ab Härnqvist (1968b). Die Stichprobengrundlage bestand aus Daten von 14000 18-jährigen Männern, die für den Militärdienst vorgesehen waren. Ungenaue Schätzungen der ersten Messungen beeinträchtigten jedoch die Ergebnisse. Gustafsson versuchte konzeptionelle Unstimmigkeiten zu umgehen und entwickelte ein hierarchisches Modell (Gustafsson, 1984). Zur Vermeidung methodischer Ungenauigkeiten konzipierte Gustafsson ein latentes Modell, das zusätzliche Faktoren wie Einkommensstufen der Personen, Schullaufbahnwahl und Eintrittsunterschiede kontrollierte. Gustafsson zieht seine Datengrundlage aus männlichen Personen, die zwischen 1972 und 1979 geboren wurden und 1996 in Schweden lebten. Ferner wurden die Daten von etwa 14000 männlichen Militärdienstankwärtlern im Alter von 18 Jahren analysiert, die eine kognitive Testbatterie zur Messung der fluiden Denkfähigkeiten (G_f), der kristallinen Intelligenz (G_c) sowie der allgemeinen kognitiven Fähigkeiten (G_v) durchführten. Gustafsson favorisiert die Messung der verschiedenen Aspekte der kognitiven Fähigkeiten. Dafür spezifizierte er ein latentes variables Modell mit Verbesserungseffekten und erwartete eine Verbesserung von drei Intelligenzquotientpunkten für die jeweiligen Schullaufbahnen. Die Ergebnisse der Studie zeigten die Veränderungen der Intelligenz anhand der individuell verschieden eingeschlagenen Schullaufbahnen während der Sekundarstufe II. Gustafsson gelang der Nachweis, dass bestimmte Schulerfahrungen Verbesserungen sowohl allgemeiner kognitiver Fähigkeiten als auch spezifischer Fähigkeiten bewirken. Auf latenter Ebene stellte Gustafsson fest, dass bestimmte Schulerfahrungen auch mit dem Abschneiden bestimmter Leistungen

in kognitiven Fähigkeitstest zusammenhängen. Gustafsson konnte konstatieren, dass Bildung die kognitiven Fähigkeiten um etwa drei IQ-Punkte verbesserte bzw. ein Zugewinn von 1,8 IQ-Punkten für jedes weitere besuchte Schuljahr zu erzielen sei. Gustafsson betont jedoch vor allem die Notwendigkeit, Ergebnisse bereichsspezifisch differenziert zu bewerten.

In der schwedischen Studie von Husén und Tuijnman war eine starke Wirkung des erfolgreichen Abschlusses hinsichtlich der kognitiven Fähigkeiten mit Abschluss der Sekundarstufe II zu verzeichnen. Husén und Tuijnman (1991) testeten zunächst die kognitiven Fähigkeiten von Personen im Alter von zehn Jahren und dann nochmals die kognitiven Fähigkeiten der gleichen Personen als erwachsene Militärdienstbewerber. Sie betrachteten die Schulabschlüsse von 2485 Personen und konnten im Mittel einen IQ-Unterschied von bis zu 2,8 Punkten feststellen. In der Studie war ein starker Zusammenhang zwischen kognitiven Fähigkeiten und dem Abschluss der Sekundarstufe II verbunden. Im Ergebnis zeigte sich ein positiver Effekt von Schulbildung in der Sekundarstufe II auf die fluiden Denkfähigkeiten und geringe Effekte auf kristallisierte Intelligenz. Eine speziell technikorientierte und wissenschaftliche Schulbildung beeinflusste dabei positiv die allgemeine Intelligenz. Der Anstieg fluiden Denkfähigkeiten wird vor allem durch die in der Schule vermittelte Argumentationsfähigkeit und die erlernte Problemlösefähigkeit zurückgeführt. Daraus schließen Husén und Tuijnman, dass Schule nicht nur oberflächliche Lerneffekte, im Sinne von Faktenwissen, erziele, sondern auch Problemlösefähigkeiten verbessert werden. Sie empfehlen, den schulischen Schwerpunkt auch auf Problemlösefähigkeiten statt lediglich auf reinen Wissenserwerb zu legen. In vergleichbaren Untersuchungen, wie der von Winship und Korenman (1997), die die Herrnstein-und-Murray-Studie analysierten, gab es einen ähnlichen geschätzten Effekt. Sie verwendeten alternative Modellspezifikationen (z. B. Vergleich von Modellen mit unterschiedlichen Annahmen über die Zuverlässigkeit der unabhängigen Variablen). Die Reanalyse führte zu höheren Schätzungen der Auswirkungen der Schulbildung auf Intelligenz. Ihre Modellspezifikation ergab einen geschätzten Effekt von 2,7 Intelligenzpunkten pro Jahr schulischer Bildung (Winship & Korenman, 1997). Sie konnten damit die ursprüngliche Position von Herrnstein und Murray nicht unterstützen, die keinen oder nur einen kleinen Effekt auf den Intelligenzquotienten feststellen. Sie sprechen sich sehr wohl für einen positiven Effekt von Bildung auf die Intelligenz aus. Aber auch Winship und Korenman lassen offen, welcher Einfluss auf die kognitiven Fähigkeiten wichtiger ist, ob Bildung oder der Einfluss früherer Intelligenzfaktoren. Brinch und Galloway

(2011) nahmen die Bildungsreform der 1960er Jahre in Norwegen, nach der der Schulbesuch von sieben auf neun Jahre verlängert wurde, zum Anlass, die Auswirkungen von Veränderungen im Bildungswesen auf die kognitiven Leistungen der betroffenen Person zu untersuchen. Die Autoren widmeten sich der einzigartigen Gelegenheit, die Veränderungen bei den betroffenen Jugendlichen im Sekundarstufenalter, die nunmehr zwei Jahre länger verpflichtend die Schule besuchen, zu verfolgen. Es war ihnen ein Anliegen, jenen Positionen, die die Wirkung von schulischen Bildungsinvestitionen auf die kognitiven Leistungen ablehnen, zu begegnen (2011). Gleichzeitig wollten sie auch jenen Positionen nachgehen, die sehr wohl der Meinung sind, dass frühe Bildung Auswirkungen auf die späteren kognitiven Fähigkeiten von Kindern hat. Brinch und Galloway (2011) konnten in ihren Untersuchungen substantielle Einflüsse schulischer Bildung auf die kognitiven Fähigkeiten von Jugendlichen im Alter von 19 Jahren finden. Im Ergebnis konnte tatsächlich eine statistisch signifikante und beträchtliche Wirkung auf die Intelligenz nachgewiesen werden. Es ließ sich, so die Autoren, ein Anstieg des Intelligenzquotienten um 0,6 Punkte für norwegische männliche Personen feststellen. Aufgrund des belegten statistisch signifikanten Anstiegs der Intelligenztestergebnisse wurden weitreichende bildungspolitische Interventionen eingeleitet. Brinch und Galloway (2011) merken jedoch an, dass die Wirkung zwischen Schulbildung und kognitiven Fähigkeiten relativ schwer bestimmbar sei, obgleich die hohe Korrelation zwischen kognitiven Fähigkeiten und Schulbesuchsdauer gut dokumentiert ist. Brinch und Galloway (2011) vertreten trotz der hohen Korrelationen die Position, dass allein aufgrund von Bildungsinvestitionen bzw. schulisch erworbenem Wissen so nicht abgeleitet werden könne, dass sie die kognitiven Leistungen bestimmen, sondern auch andere Faktoren wirken. Damit schließen sich Brinch und Galloway (2011) teilweise jenen Positionen an, die gegenseitige oder multiple Faktoren auf die kognitiven Leistungsmaße annehmen. Sie empfehlen daher Untersuchungen von Interventionen in der frühen Kindheit sowie der Umweltbedingungen, um die Entwicklung kognitiver Fähigkeiten genauer zu analysieren. Aber ein wesentlicher Befund ist, dass selbst relativ späte Interventionen in der Adoleszenz auf eine bestimmte Ursache (hier die Bildungsinvestitionen, Verlängerung um zwei Jahre) zurückgeführt werden können. Nachgewiesen werden konnte auch, dass sich eine späte Einschulung oder unregelmäßiger Schulbesuch nachteilig auf die intellektuellen Fähigkeiten von Kindern im Vergleich zu Gleichaltrigen, die regelmäßig die Schule besuchen, auswirken (Sherman und Key, 1932, in Brinch & Galloway, 2011) befassten sich erstmals mit den Zusammenhängen zwischen Schulbesuch und Intelligenzleistungen von Kin-

dern. Sie beobachteten, dass die Intelligenzleistungen von sogenannten Höhlenkindern systematisch mit der Verweildauer dort korrelierten. Es zeigten sich auch dramatische altersbedingte Zusammenhänge. Je älter das Kind, desto niedriger war sein Intelligenzquotient im Vergleich mit gleichaltrigen Kindern, die die Schule besuchten. Sechsjährige Kinder hatten einen vergleichsweise kleinen Intelligenzunterschied zu gleichaltrigen Kindern, die die Schule besuchten, während Kinder im Alter von 14 Jahren deutliche negative Abweichungen aufwiesen. Freemann (1934, in Brinch & Galloway, 2011) untersuchte zwei Jahre später die Wirkung von Bildungswissen auf die Intelligenz von Kindern mit niedrigen Intelligenzwerten. Einige dieser Kinder hatten die Schule nur unregelmäßig besucht. Es handelte sich hauptsächlich um Kinder, deren Eltern Binnenschiffer waren und ihre Kinder auf ihre Fahrten mitnahmen. Sie stellte fest, dass diese Kinder ein vergleichsweise niedrigeres Intelligenzniveau hatten als Kinder, die die Schule regelmäßig besucht hatten. Auch bei DeGroot (1951) stand der unterbrochene Schulbesuch von Kindern, die aufgrund von regionalen Besonderheiten die Schule nur zeitweise besuchen konnten und denen der Zugang zur Schulbildung erschwert wurde, bzw. Kinder, die aufgrund von Kriegswirren die schulische Ausbildung verspätet beginnen konnten, im Zentrum der Forschungsbemühungen. Kinder, die den Schulbesuch für längere Zeit unterbrechen mussten, hatten ein bis zu sieben Intelligenzwertpunkte niedrigeres Intelligenzniveau im Vergleich zu altersgleichen Kinder mit regelmäßigem Schulbesuch. Von ähnlich hohen kumulativen Defiziten der kognitiven Fähigkeiten im Zusammenhang mit dem Schulbesuch berichteten auch Vernon (1969), der in Südafrika die Auswirkungen eines verzögerten Schulbeginns untersuchte. Kinder, deren Schulbesuch wegen Nichtverfügbarkeit von Schulen für bis zu vier Jahre verzögert wurde, hatten einen niedrigeren Intelligenzquotienten von fünf Intelligenzpunkten für jedes Jahr, um das ihre Beschulung verzögert wurde. Die Ergebnisse wurden insgesamt auf korrelativer Basis ermittelt. Doch nicht nur die Schulbesuchsdauer stand im Mittelpunkt bisheriger Forschungsbemühungen, sondern auch der intermittierende Schulbesuch oder die Unterbrechung der Schulausbildung während der Sommerferien war Gegenstand diverser Untersuchungen. In diesem Zusammenhang ist es interessant, dass selbst eine relative kurze Unterbrechung des Schulbesuchs zum Beispiel in den Sommerferien Auswirkungen auf die kognitiven Fähigkeiten zeitigt. Eine Erklärung hierfür liefert das Intelligenzmodell von Cattell (1957, 1963, 1965, 1971, 1975). Das in der Schule erworbene Wissen könne demnach den kristallisierten Intelligenzanteilen zugeordnet werden. Es gerate zum Teil in den Ferien in Vergessenheit, da einige Inhalte im Alltag kaum benötigt werden. Bekannt ist in diesem Zusammenhang vor allem die Sommerferienstudie von Jencks et al. (1972). Sie stellten fest,

dass bei Kindern während der Sommerferien eine Abnahme des Intelligenzquotienten zu verzeichnen war. Die Ferien werden insofern als eine kognitive Trainingspause betrachtet, deren Einfluss messbar wird. Diese Abnahme war besonders stark bei Kindern mit niedrigem sozioökonomischem Hintergrund, woraus geschlossen wurde, dass Kinder aus akademischen Elternhäusern auch in den Sommerferien Bildungsangebote nutzen bzw. anderweitige Bildungsangebote diese Abnahme mindern. Auch Cooper et al. (1996) untersuchten in einer Metaanalyse über 39 Studien den Einfluss der Sommerferien auf die Leistungen in Mathematik, Lesen und Rechtschreibung. Sie konnten feststellen, dass die Sommerferien etwa zu einem durchschnittlichen Verlust eines Monats einer Klassenstufe führen. Dabei wirkte die Sommerpause sich nachteiliger auf die Leistungen in Mathematik als auf die Leistungen im Lesen oder in der Rechtschreibung aus. Die negativen Auswirkungen nahmen mit dem Anstieg des Schülerniveaus zu. Die Autoren interpretieren den geringeren Verlust der Deutschleistungen mit der Verfügbarkeit von mehr Lesepraxis in den Ferien und den unterschiedlichen Möglichkeiten, die erlernten Fähigkeiten in der Praxis weiterhin zu nutzen. Erworbene mathematische Fähigkeiten sind darüber hinaus anfälliger für den Gedächtnisverfall (Fakten und verfahrensbezogene Kenntnisse werden leichter vergessen als begriffliches Wissen) und auch in den Ferien vermutlich weniger genutzt. Als pädagogische Intervention schlugen die Autoren vor, Sommerschulprogramme zu initiieren oder auch Schulkalenderänderungen vorzunehmen. Hinsichtlich der sozioökonomischen Unterschiede untersuchten Alexander et al. (2001), ob die Leistungsdifferenz mit der sozialen Schichtung von Grundschulkindern zusammenhängt. Anhand einer repräsentativen Datengrundlage von Schulkindern aus Baltimore konnten sie zeigen, dass die außerunterrichtliche Anregung während der Ferien Grund für einen Leistungszuwachs bestimmter Kinder während der Sommerferien ist. Die Effekte waren geringer als die Effekte, die während des laufenden Schuljahres erzielt werden, doch die Autoren betonen, dass die Schulbildung selbst anscheinend eine wichtige Kompensationsrolle für sozial benachteiligte Kinder spielt. Campione und Brown (1978) führten vergleichende Studien zwischen retardierten und nicht retardierten Kindern durch. Dabei stellten sie sich die Frage, inwieweit kognitive Fähigkeiten durch Schulleistungen bzw. expliziter durch Vorwissen beeinflusst werden können und durch Trainingsmaßnahmen beeinflussbar seien. Sie konnten dies bestätigen. Ähnlich argumentieren auch Cahen und Cohen (1989) und Ceci (1991), welche den Trainingsanteilen einen bedeutenden Einfluss auf Gedächtnisleistungen zuschreiben. Diese Autoren merken zudem an, dass in einer Vielzahl gefundener Studien eine problematische Unterscheidung

zwischen den Begriffen Intelligenz und schulischer Leistung vorgenommen werde. So beschäftigen sich Cahen und Cohen auch (1989) mit dem Einfluss der formalen Bildung auf die Intelligenz. Sie wenden ein, dass viele vorliegende Studien zwar das gleiche Alter von Schülern annehmen, doch Differenzen innerhalb der Schulbildung oder auch anderer in Beziehung zur Intelligenz stehender Variablen außer Acht lassen. Die Schwierigkeit bestehe darin, Gleiches mit Gleichem zu vergleichen. Sie verweisen in diesem Zusammenhang auf kulturelle Unterschiede oder Unterschiede im Schulsystem bzw. soziale Unterschiede. Die Autoren resümieren aufgrund ihrer Untersuchungen anhand von Daten von Grundschulern, dass der Schulbesuch einen großen Einfluss auf die Intelligenzentwicklung hat. Sie interpretieren im Sinne Cattells, dass die Personen mit der meisten Schulbildung auch über eine breitere kristallisierte Intelligenz verfügen. Sie berichten, dass die Auswirkung der Schulbesuchsdauer auf die Steigerung der Intelligenztestergebnisse im Vergleich mit dem chronologischen Alter etwa doppelt so stark war. Sie fanden eine beträchtliche Wirkung der Schulbildung. Diese bisher vorgestellten Ergebnisse werden zumeist als ein Beweis für den kausalen Zusammenhang zwischen Schulbildung und Intelligenz interpretiert. Es zeigen sich systematische Veränderungen der Intelligenzscores im Zusammenhang mit der Dauer der (Vor-)Schulbildung, die ein Kind erhält, doch machen diese Studien insgesamt recht unpräzise Aussagen zu den Zusammenhängen zwischen Bildung und Intelligenz.

Zahlreiche andere Forschungen, die sich mit den Zusammenhängen zwischen kognitiven Fähigkeiten und Schulleistungen befassen, versuchen im Gegensatz zu der oben dargelegten Position, die dominante Rolle der kognitiven Fähigkeiten zu belegen und beziehen sich dabei auch auf Binet. Aus der Perspektive Binets kann erwartet werden, dass Intelligenzleistungen einer Person deren Schulerfolg vorhersagen können. Historisch betrachtet wurden Intelligenztests also entwickelt, um die Fähigkeit der Schüler in der Schule zu messen und nicht umgekehrt. Diese Annahme wird seit Beginn der Intelligenzforschung empirisch unterstützt (Kamphaus, 2001). Intelligenzmessungen dienen traditionell der Fähigkeitsschätzung von Schülern hinsichtlich des Schulerfolgs. Binet und Simon entwickelten Intelligenztests unter dieser Annahme und konnten die Zusammenhänge deutlich machen (Binet & Simon, 1916). Auch Crano, Kenny und Campbell (1972) widmeten sich der Beziehung zwischen den Konstrukten kognitive Fähigkeiten und Schulleistungen. Sie versuchten bereits 1972 die Beziehung im Längsschnitt (1963-1967) zu untersuchen. 5495 Schüler der vierten Klassenstufe absolvierten einen Intelligenztest, sowie einen Test zur Erfassung von Bildungsleistungen. Zwei Jahre später wurden diese Schüler erneut mit dem

Testverfahren konfrontiert. Crano, Kenny und Campbell (1972) fragten, ob eher die Übernahme von spezifischen Fähigkeiten oder das Erlernen von spezifischen Informationen (Leistung) zu einer erhöhten Fähigkeit zur Abstraktion im Sinne von Intelligenz führt oder die vorhandene Intelligenz die Leistung verursacht. Die Autoren sprechen sich aufgrund ihrer Ergebnisse für eine höhere Prognosekraft von Bildungsleistungen aufgrund früherer kognitiver Fähigkeiten aus. Diese Studie wurde später von Watkins (2007) hinsichtlich ihrer statistischen Gründlichkeit untersucht, von ihm jedoch nur als eingeschränkt haltbar bewertet (Watkins, 2007). Statistische Ungenauigkeiten und eingeschränkte Stichprobenauswahl sowie die Auswahl der Testverfahren würden den Gütekriterien kaum genügen. Auch Messfehler beeinträchtigten die Analysen, so Watkins. So wurden teils städtische und ländliche Stichproben miteinander verglichen. Zudem stützten sich die Daten auf nur einen Schulbezirk und viele Intelligenztestergebnisse wurden lediglich auf der Gesamtscore-Ebene analysiert. Rückschlüsse dieser Ergebnisse auf aktuelle Zusammenhänge zwischen den Konstrukten seien, so Watkins (2007), nicht angemessen, da die Daten bereits vor mehr als 50 Jahren erhoben (1963-1967) und Tests eingesetzt wurden, die 1957 konzipiert wurden.

Die psychologische Forschung stellt die Bedeutung der kognitiven Fähigkeiten für den Leistungskontext heraus und sieht vornehmlich kognitive Voraussetzungen als Disposition zur Leistungserbringung (Stern, 1997, 2004; Helmke 1997, 2008; Heller, 1991). Die Studie von Fergusson, Horwood und Ridder (2005) bestätigte erneut den Zusammenhang zwischen kognitiven Fähigkeiten im Alter von sieben und acht Jahren und wissenschaftlichen Leistungen, die diese Personen im Erwachsenenalter vollbringen, unabhängig von kindlichem Verhalten sowie familiären und sozialen Verhältnissen (Fergusson, Horwood, & Ridder, 2005). Allerdings beruht diese Untersuchung auf korrelativen Analysen. Auch auf korrelativer Ebene belegten Colom und Flores-Medosa (2007) den Zusammenhang zwischen Intelligenz und schulischen Leistungen. Sie konnten anhand einer brasilianischen Stichprobe von 641 Kindern die Zusammenhänge zwischen Intelligenzleistung, sozioökonomischen Faktoren wie Einkommen und Bildungsstand der Eltern und schulischen Leistungen untersuchen. Die Ergebnisse deuten darauf hin, dass die sozioökonomischen Faktoren die Schulleistungen nicht vorherzusagen vermochten, während die kognitiven Fähigkeiten sehr wohl die schulischen Leistungen vorhersagen konnten. Damit unterstreichen diese Ergebnisse die Bedeutung der kognitiven Fähigkeiten als Prädiktor schulischer Leistungen. Unter den wenigen existierenden Studien für das Grundschulalter standen wegen ihrer pädagogischen Relevanz am häufigsten die Zusammenhänge zwischen kognitiven Merkmalen und mathematischen Schulleistungen im Zentrum von Forschungsbemühungen,

da letztere als bedeutendster Prädiktor für schulischen Erfolg gelten (Weber, 2013, 2015; Helmke, 1997). Den Einfluss von kognitiven Fähigkeiten auf mathematische Schulleistungen betonte beispielsweise Stanovich (1986). Werden Zusammenhänge auf korrelativer Ebene analysiert, so wird davon ausgegangen, dass die Korrelationen zwischen Intelligenz und Mathematikleistungen zwischen .39 bis .62 liegen (Stern, 1997, 2004). Von einer Beeinflussung von Mathematikleistungen durch die Intelligenz gehen Lubinski und Dawis (1992) aus. Doch den eindeutigen Nachweis zu erbringen, ist auch deshalb schwierig, da die Ursachen aktueller Beobachtungen in der Vergangenheit zu suchen sind und vormalig das möglicherweise aktive Aufsuchen von anregenden Bedingungen, wie Ziegler in seinem Aktiotopmodell beschreibt (2005), eine bedeutende Rolle auf die kognitive Fähigkeitsentwicklung spielten. Das Aktiotopmodell beschreibt die für eine Weiterentwicklung notwendige Veränderung der Lernumgebung, die von der Person eigens initiiert wird und andere Personen anregt, eine für die betreffende Person geeignete Umgebung herbeizuführen. Dabei ist der unbestimmte Einfluss von Messfehlern häufig Ursache für die fehlende Beweislage. Insgesamt ist die retrospektive Analyse der Zusammenhänge ohnehin problematisch. Auch bei Floyd et al. (2003) stand der Zusammenhang zwischen fluiden Denkfähigkeiten im Sinne Cattells (1971) und Mathematikleistungen im Mittelpunkt der Untersuchungen (Floyd et al., 2003). Die Autoren betonen den engen Zusammenhang der flüssigen Intelligenz mit der mathematischen Leistung und unterstreichen mit ihren Ergebnissen den dominanten Anteil fluider Denkfähigkeiten bei der Bestimmung des Mathematikleistungsniiveaus. Rohde und Thompson widmeten sich (2007) in ihrer Studie ebenfalls dem Anliegen, akademische Leistungen anhand kognitiver Fähigkeiten und spezifischer kognitiver Fähigkeiten zu erklären. Dabei setzten sie zwei Leistungstests zur Erfassung akademischer Leistung ein. Zur Erfassung der kognitiven Fähigkeiten wurden die Matrizen-Testverfahren zur Messung allgemeiner kognitiver Fähigkeiten und ein Vokabeltest angewendet. Sie maßen Fähigkeiten wie Arbeitsspeicher, Verarbeitungsgeschwindigkeit sowie räumliche Vorstellungsfähigkeit und weiterhin die allgemeine kognitive Fähigkeit anhand in einer Stichprobengröße von 71 erwachsenen Personen. Es wurden jedoch keine spezifischen kognitiven Fähigkeiten erfasst. In der leider nur kleinen Stichprobe konnten sie dennoch nachweisen, dass allgemeine Fähigkeiten akademische Leistungen vorherzusagen vermögen. Sie konstatierten Zusammenhänge zwischen .50 bis .75 zwischen Intelligenzleistungen und akademischen Leistungen. Jedoch konnten keine spezifischen kognitiven Fähigkeiten prognostiziert werden. Im Ergebnis bildeten die Verarbeitungsgeschwindigkeit und das räumliche Vorstellungsvermögen signifikante Einflussgrößen bei der Vorhersage von mathematischen

Leistungen (festgestellt mit dem SAT-Test). Rohde und Thompson (2007) thematisieren schlussfolgerndes Denkvermögen als besonders relevanten Prädiktor für das Leistungsniveau in naturwissenschaftlichen Fächern, während es für sprachlich orientierte Schulleistungen weniger relevant ist. Krumm et al. (2008) konnten ebenso zeigen, dass kognitive Fähigkeiten mit der Mathematikleistung stärker assoziiert werden als mit sprachlichen Domänen. Sie untersuchten den Einfluss von schlussfolgerndem Denken und Arbeitsgedächtniskapazität als Prädiktoren von Schulleistungen und stellten fest, dass schlussfolgerndes Denkvermögen als bester Prädiktor der Schulleistungen fungiert. Allein der belegte Zusammenhang zwischen den Konstrukten Intelligenz und schulischen Leistungen schließt jedoch laut Weber et al. (2013) nicht den Einfluss auch anderer Faktoren aus. Dies nahmen sie zum Anlass, den Einfluss anderer potentieller möglicher Variablen auf den Zusammenhang zu untersuchen. Im Rahmen ihrer Untersuchung, welche die Rolle von kognitiven und motivationalen Prädiktoren für die Schulleistungen in Mathematik und Deutsch untersuchen sollte, absolvierten 320 deutsche Grundschulkinder der vierten Klassenstufe einen Intelligenztest, einen Test der Arbeitsspeicherkapazität und ein Motivationstestverfahren. Die Analysen erfolgten anhand von Strukturgleichungsmodellen. Weber et al. konnten zeigen, dass die Motivation als Prädiktor eine erhebliche Varianz der Abweichungen domänenspezifischer Schulnoten erklären konnte. Ein wesentlicher Befund jedoch war, dass sich kognitive Fähigkeiten wiederum als stärkster Prädiktor der Mathematikleistungen erwiesen (COG: $\beta = .59$; MOT: $\beta = .41$). Im Gegensatz dazu wirkte die Motivation für das Fach Deutsch stärker als der kognitive Prädiktor (COG: $\beta = .34$; MOT: $\beta = .67$). Motivation und Intelligenz gemeinsam konnten 75 % der Mathematikleistungen vorhersagen, während der Einfluss der Intelligenz in dem Unterrichtsfach Deutsch geringer war, so Weber et al. (2013). Dabei verwiesen Weber et al. auf den dominierenden Einfluss der Intelligenz und betonten die kognitiven Fähigkeiten als wichtigste Prädiktoren mathematischer Leistung und des Erfolges (Weber et al. 2015). Die Aussagequalität der Untersuchung wird allerdings eingeschränkt durch die Tatsache, dass zur Messung der fachbezogenen Leistungen lediglich Schulnoten erfasst wurden und keine Schulleistungstests, die schulische Leistungen valider und objektiver erfassen, wie unter anderen Hesse und Latzko (2011) und auch Ingenkamp und Lissmann (2008) betonen. Zudem wurden die kognitiven Fähigkeiten im Sinne der allgemeinen kognitiven Fähigkeit g erfasst, sodass keine Einflüsse bereichsspezifischer kognitiver Fähigkeiten und spezifischer schulischer Leistungen erfasst wurden. Damit sind keine Rückschlüsse über spezifische Wirkungen erlaubt. Weiter wurde nur eine relativ geringe Stichprobengröße untersucht und die hierarchische Datenstruktur der teilweise verschachtelten Datengrundlage

konnte nicht berücksichtigt werden, so konstatieren Weber et al. (2013) selbst. Eine weitere Studie, die Lernen und Wissenserwerbsprozesse hinsichtlich ihrer Beeinflussung durch kognitive Faktoren in den Mittelpunkt stellte, wurde ebenso von Weber et al. (2015) durchgeführt. Sie untersuchten wiederum den Zusammenhang zwischen den kognitiven Fähigkeiten und Mathematikleistungen unter dem Aspekt des Lernverhaltens und stellten fest, dass für Leistungen im Fach Deutsch vor allem das Lernverhalten verantwortlich sei, während sich der Gesamtintelligenzquotient erneut als alleiniger Prädiktor der Mathematikleistungen erwies. Die Autoren schlussfolgerten, eine gute Mathematikleistung erfordere demnach im Gegensatz zu den Anforderungen im Fach Deutsch eine Vielzahl kognitiver Fähigkeiten, wie Problemlösefähigkeiten und schlussfolgerndes Denken. Allerdings basierte auch diese Untersuchung von Weber et al. (2015) auf einer Datenerhebung von lediglich 50 Schülern, die in Jugendhilfeeinrichtungen betreut wurden. Eine weitere Kontrollstichprobe von 45 Schülerinnen und Schülern absolvierte ebenfalls einen nonverbalen Test zur Messung kognitiver Fähigkeiten. Die Mathematikleistungen wurden erneut über Schulnoten erhoben. Auch das Alter der Teilnehmer der vorliegenden Stichprobe blieb unberücksichtigt und Mathematik- und Deutschnoten waren nicht normal verteilt, denn 74 % der Experimentalgruppe hatten einen Gesamtintelligenzquotienten im unteren Normbereich oder darunter. Aussagen über die Kausalbeziehungen zwischen kognitiven Fähigkeiten und Mathematiknoten sind somit erneut nicht möglich (Weber et al. 2015). In der Studie von Karbach et al. (2013) wurde ebenfalls festgestellt, dass die kognitiven Fähigkeiten großen Einfluss auf die Mathematikleistungen und weniger auf sprachliche Domänen haben. Karbach et al. untersuchten den Einfluss der elterlichen Unterstützung im Zusammenhang mit kognitiven Fähigkeiten auf die schulische Leistung im frühen Jugendalter mittels latenter Strukturgleichungsmodelle. Das Alter der 334 Kinder betrug durchschnittlich 12,4 Jahre (10 bis 14 Jahre). Dabei stand auch der Einfluss der kognitiven Fähigkeiten im Mittelpunkt. Diese – so konnten die Autoren feststellen – fungierten als Hauptfaktor der schulischen Leistung. Nach der Kontrolle der Intelligenz erwies sich jedoch die elterliche Unterstützung als sehr signifikant für die kindliche Leistungsentwicklung in der Schule. Die Vorhersage von Mathematikleistungen aufgrund der generellen kognitiven Fähigkeiten gelang mit .54, gemessen mit verbalen und nonverbalen Fähigkeitstests. Es wurde auch deutlich, dass die elterliche Bildung einen indirekten Einfluss auf die kognitiven Fähigkeiten hat (Karbach et al., 2013). Jäger, Süß und Beauducel (1997), Süß (1996) wollten der Tatsache Rechnung tragen,

dass einmalige Messungen präzise Aussagen zu dem zeitlichen Verlauf der Beziehung zwischen den Konstrukten kognitive Fähigkeiten und mathematische Schulleistungen nicht zulassen.

So widmeten sich Jäger, Süß und Beauducel (1997), Süß (1996, 2003) und Brunner (2006) und auch Süß und Beauducel (2015) explizit der Frage nach der Einflussrichtung zwischen den Konstrukten. Die Autoren wiesen die dominierende Rolle kognitiver Fähigkeiten für Leistungen im Fach Mathematik nach und belegten die Abhängigkeit des Schulleistungsverlaufs von früherem kognitiven Leistungspotential. Auch Primi et al. (2010) fokussierten ihre Aufmerksamkeit auf die Entwicklungsverläufe. Die Niveauveränderungen der Mathematikleistungen in Abhängigkeit von den kognitiven Fähigkeiten standen in ihrer Untersuchung im Zentrum. Sie definieren die kognitiven Fähigkeiten als vornehmlich fluide Intelligenz, die als allgemeine Fähigkeit zu lernen eingeschätzt wird. Dabei setzen sie voraus, dass höhere Intelligenz auch eine höhere Lernfähigkeit beinhaltet. Sie untersuchten den Zusammenhang zwischen fluider Intelligenz auf Mathematikleistungen anhand einer Stichprobe von 166 portugiesischen Mittelstufenschülern im Alter von 11 bis 14 Jahren. Diese absolvierten vier Mathematikleistungstests jeweils zu Beginn und am Ende der siebten und achten Klassenstufe. Kognitive und mathematische Testverfahren wurden auf Subtestebene erhoben. Die Analysen erfolgten anhand mehrstufiger Wachstumskurvenmodellierungen. Die Ergebnisse zeigen, dass ein hohes Intelligenzmaß zu Beginn auch mit einem steileren Anstieg der Mathematikleistungsscores verbunden war. Die Mathematikleistungstests, die eingesetzt wurden, deckten das Kerncurriculum der jeweiligen Klassenstufe ab. Sie erfassten Anteile aus den Bereichen Geometrie, Gleichungssysteme, Statistik, mathematische Funktionen und Arithmetik. Die Intelligenztests wiederum basierten auf der Messung der Fähigkeiten zur Bildung von Analogien, zu Herstellung kognitiver Beziehungen, ferner auf der Erfassung quantitativen Denkvermögens, der Fähigkeit zu verbalen Schlussfolgerungen und der Messung von räumlichem Vorstellungsvermögen. Die Ergebnisse zeigten zunächst eine große Variabilität der intraindividuellen Wachstumsentwicklung, die mit der Intelligenz im Zusammenhang steht, da alle Korrelationen positiv sind. Weiterhin zeigen die Ergebnisse einen Zusammenhang zwischen der Lerngeschwindigkeit im Sinne der flüssigen Intelligenz und den mathematischen Ergebnissen. Die starke Beziehung zwischen fluiden Intelligenzanteilen und dem numerischen Schlussfolgern sowie den Mathematikleistungen wurde sehr deutlich. Anfängliche fluide Denkfähigkeiten dienten dabei als bedeutender Prädiktor für die Wachstumsrate der Mathematikleistungen. Die Autoren unterschieden im Anschluss ihrer ersten Untersuchungsergebnisse ferner die kognitiven Gruppierungen nach

niedrigem Niveau, durchschnittlichem Niveau und höherem Niveau. Der Vergleich zeigte eine weniger steile Entwicklung für die Gruppe mit niedrigem Intelligenzniveau. So konnten schließlich die Autoren ihre Anfangshypothese bestätigen, nach der fluide Intelligenz die Wachstumsrate der Mathematikleistungen vorhersagen kann. (Primi et al., 2010) schlussfolgern, dass kognitive Fähigkeiten nicht nur das Anfangsniveau mathematischer Leistungen, sondern auch die Lernleistung im Laufe der Zeit bestimmen. Voelke et al. (2006) analysierten in der Reanalyse einer Studie von Ackerman, Kanfer und Goff (1995 in Voelke, et al. 2006) die Beziehung zwischen fluider Intelligenz und einer Lernaufgabe für Flugverkehrsleiter. Dabei konnten auch sie die fluide Intelligenz als wesentlichen Faktor des Lernens ermitteln. Damit stehen sie der Position von Luo et al. (2006) entgegen, die dafür plädieren, die Testverfahren auf die Messung kristalliner Intelligenzanteile auszurichten und Testverfahren, die fluide Fähigkeiten erfassen, einer kritischen Prüfung zu unterziehen. Vor allem in neuartigen Situationen unterstützt diese jedoch die Aneignung von neuen Fähigkeiten und Kenntnissen. Dabei resümieren sie, dass vor allem kognitive Fähigkeiten die Lernleistungen bestimmen. Voelke et al. betonen anhand einer Wachstumskurvenmodellierung den engen Zusammenhang zwischen Fähigkeiten und Lernleistung. Um ein besseres Verständnis der Veränderungen von Leistungsstrukturen zu erreichen, sei es vorteilhaft, neue methodische Ansätze zu etablieren. Sie können – im Gegensatz zu der reinen Korrelationsanalyse – komplexe, zeitversetzte und auch kovariante Beziehungen sowie Fehlerstrukturen integrieren. Sie belegten, dass eine präzise Modellspezifikation wesentlich zur Aufklärung der Zusammenhänge beitragen kann und plädieren für die Verwendung von theoretischen und statistisch fortschrittlichen Techniken der Analyse zur Messung von Veränderung. Während die bisher vorgestellten Untersuchungen die kognitiven Fähigkeiten als dominierenden Faktor betrachten, finden andere Forscher, die diese Ansicht im Grundsatz teilen, trotzdem auch weitere Einflussfaktoren. Yen, Konold, und McDermott (2004) untersuchten beispielsweise den Einfluss von Lernverhalten in einer Mehrgruppenanalyse mit einer Stichprobe von 1304 Schülern. Im Ergebnis erwiesen sich zwar die kognitiven Fähigkeiten als dominanter Prädiktor schulischer Leistungen, doch der Einfluss des Lernverhaltens sei nicht von der Hand zu weisen. Daher sei es pädagogisch sinnvoll, den Einfluss des Lernverhaltens in pädagogischen Situationen zu berücksichtigen.

Laidra et al. (2007) untersuchten aus einem Datenbankbestand von isländischen Grundschulkindern verschiedener Klassenstufen die allgemeine Intelligenz mit den Raven Standard Progressive Matrices. Daneben wurden Persönlichkeitsmerkmale mit dem Erhe-

bungsinstrument NEO Five Factor Inventory und Schulleistungen erhoben. Die erneute Untersuchung erfolgte in den Klassenstufen 6 bis 12. Die Gewissenhaftigkeit nahm unter den Persönlichkeitsfaktoren eine besondere Bedeutung ein, dennoch erwies sich die Intelligenz als stärkster Prädiktor der Schulleistungen. Dieses Ergebnis zeigt zum einen den Einfluss der kognitiven Fähigkeiten auf die späteren Schulleistungen, wenngleich hier sehr große Zeitspannen zwischen den Erhebungen liegen. Es zeigt aber auch den Einfluss weiterer Faktoren, die im Einklang mit den Theorien von Gagné (2004); Perleth, (1997), Perleth und Sierwald (1992); Ceci (1991) stehen. Nicht zuletzt stellten Deary et al. (2007) – wie bereits oben erwähnt – durchaus auch den Einfluss anderer Faktoren auf die Entwicklung der kognitiven Fähigkeiten fest, als sie in einer 5-Jahres-Längsschnittstudie mit 70000 eine .81-Korrelation zwischen Intelligenz im Alter von 11 Jahren und Bildungsleistungen im Alter von 16 in 25 akademischen Fächern belegen konnten. Mackintosh (2011) berichtet zwar ebenfalls von Korrelationen in Höhe von .50. Doch er betont, dass die Beziehungen zwischen einem Intelligenzquotienten im Alter von zehn Jahren und einem späteren Intelligenzquotienten der Person im Alter von 25 oder 40 Jahren sehr komplex sei. So werde die weitere Entwicklung auch von den Umweltbedingungen bestimmt. Dies verdeutlicht er am Beispiel junger brasilianischer Straßenverkäufer, die komplexe Arithmetikkenntnisse im Arbeitskontext des Melonenverkaufs erworben haben. Mackintosh verweist weiter auf den großen Einfluss von Erfahrungen und Trainingsintensität, die im Laufe der Zeit eine Person zu einem Experten machen. In dem Zusammenhang verweist er auf die Ergebnisse aus der Expertiseforschung. Extreme Positionen der Expertisetheorie lehnen den Einfluss kognitiver Fähigkeiten bei der Leistungserbringung sogar vollkommen ab und betonen stattdessen den Einfluss von Training, Übung und Motivation bei der intensiven Auseinandersetzung mit einem thematischen Bereich (Ericsson et al., 1993). Sie gehen davon aus, dass die Leistungsqualität in Abhängigkeit der Beschäftigungszeit mit dem Fachgebiet und damit der Intensität der Übung steht (Ericsson, Krampe und Heizmann, 1993). Dabei seien vor allem die Qualität der Übungsprozesse und vor allem der Einfluss von Erfahrungen und Lernprozessen entscheidend. Unter Erfahrung wird ein in bereichsspezifischen Lernprozessen erworbenes deklaratives und prozedurales oder auch strategisches Wissen verstanden. Eine umfangreiche Wissensbasis bildet die Grundlage herausragender Leistungen auf verschiedenen Gebieten. Mathematisches Schulwissen sei demnach bereichsspezifisches Wissen, das auch in diesem Kontext entsprechend umfangreich bei hinreichender Auseinandersetzung mit dem Gebiet erworben werden könne. Lernprozesse gelängen zudem umso erfolgreicher, je kompetenter, motivierter und erfahrener die Lehrkräfte sind, die ihn begleiten

(Ericsson, Krampe & Tesch-Römer, 1993d). Diese Vorstellungen gingen in das Begabungsmodell von Heller (1992c) und Perleth und Heller (1994) und in die Modellerweiterung von Perleth (1997) ein. Letztlich fließen in die theoretischen Überlegungen ein, dass die intensive Auseinandersetzung mit einem Bereich auf die kognitive Leistung wirkt (Perleth, 1997). Das Hauptproblem der Expertisevertreter sei jedoch, so die Kritiker, dass die Beurteilung des Erwerbs von Expertise erst in der retrospektiven Perspektive möglich ist. Also erst die Befragung von Experten, wie sie zum Experten geworden sind, lässt Rückschlüsse zu. Den Expertiseerwerb zuvor zu prognostizieren, ist empirisch nicht gelungen. Aus der retrospektiven Perspektive werden aber seitens der Begabungsforschung zumeist eher kognitive Voraussetzungen für herausragende gezeigte Leistungen diskutiert (Heller, 1991, 1997). Helmke (1997) beschreibt in dem „Modell der Bedingungsfaktoren schulischer Leistungen“ (in Helmke & Weinert, 1997) auf der übergeordneten Ebene die Schulorganisation und schulische Einflüsse als Moderatoren schulischer Leistungen. Auf der anderen Seite wirken sozioökonomische und familiäre Strukturen auf die Persönlichkeit der Eltern und über Erziehungsfaktoren wiederum auf die Persönlichkeit des Kindes. Diese wird ihrerseits auch von kognitiven und affektiven Merkmalen bestimmt. In der SCHOLASTIK-Studie wurden kognitive Variablen im Sinne allgemeiner Intelligenz, Schulleistungen, Schulnoten und motivationale Faktoren erhoben. Die intellektuellen Fähigkeiten wurden in der ersten Klassenstufe und in der vierten Klassenstufe erfasst. Es wurden die Zusammenhänge zwischen Schulnoten und Mathematikleistungstests ermittelt, die bei .63 in der zweiten Klassenstufe lagen und bei .76 in der dritten Klassenstufe, sowie in der vierten Klassenstufe bei .72. Damit sind die Korrelation zwischen Mathematikleistungstestergebnissen und Schulnoten relativ hoch, was im Gegensatz zu den Ergebnissen von Sauer und Gamsjäger (1996) steht, die den Schulnoten eine relativ geringe Validität zugesprochen haben. Im Rahmen der Untersuchung zeigte sich eine Korrelation zwischen Mathematikleistungen der zweiten Klassenstufe und dem bereichsspezifischen Vorwissen in der dritten Klassenstufe von .55 und in der vierten Klassenstufe in Höhe von .84. Dem gegenüber stehen die Korrelationen zwischen CFT 20 und den Mathematikleistungen, die eine Höhe von .51 in der zweiten Klassenstufe haben und eine Korrelationen von .47 in der dritten und vierten Klassenstufe zeigen. Helmke nennt die Intelligenz erwartungsgemäß als stärksten Prädiktor. Die kognitive Fähigkeit erklärt 41,8 % der Mathematiktestleistung, im Zusammenhang mit der Motivation und den kognitiven Fähigkeiten werden 29,6 % erklärt. Unerklärt bleiben 20,8 %. In der Betrachtung der Mathematiknote erklären kognitive Fähigkeiten 22 % und kognitive Fähigkeiten und Motivation gemeinsam 30,7%. Unerklärt bleiben 41,1%. Damit stellt Helmke

den überragenden Stellenwert kognitiver Fähigkeiten heraus (Helmke, 1997). Heller (1991) positioniert sich 1991 mit der Aussage, dass im Rahmen der Vorhersage von Schulerfolg bzw. Schulleistungen von multifaktorieller Beeinflussung auszugehen sei. Obwohl auch er kognitive Fähigkeiten zur Bestimmung des Schulerfolges als dominant ansieht, spaltet er in seinem Schulleistungsbedingungsmodell die Prädiktorvariablen in Faktoren der Schülerpersönlichkeit und Faktoren der sozialen Umwelt auf. Auf Seiten der Schülerpersönlichkeit nennt er Vorwissen, kognitive Fähigkeiten und nicht-kognitive Merkmale. Auf der anderen Seite der sozialen Lernumwelt werden als Prädiktorvariablen Schule, Familie und Einfluss der Gleichaltrigen genannt. Im Rahmen diagnostischer Prozesse sieht er für die Entscheidungsfindung die Ergebnisse aus Fähigkeitstests und Schulleistungen als maßgeblich, wie er in seinem Modell „Schuleignungsermittlung auf der Basis einer sequentiellen Entscheidungsstrategie“ beschreibt. Heller (1997) betont die Bedeutung der Wechselwirkung zwischen kognitiven und motivationalen Faktoren und der sozialen Umgebung. Erneut weist er jedoch darauf hin, dass den kognitiven Fähigkeiten und dem Vorwissen sowie der Motivation, aber auch dem schulischen Unterricht eine wesentliche Bedeutung zukommt. Zugleich macht er deutlich, dass die Schullaufbahnberatung mittels kognitiver Prädiktoren unverzichtbar sei. Trotz der oben berichteten Zusammenhänge bleibt es dennoch schwierig, einen Nachweis zu erbringen, da die Ursachen aktueller Beobachtungen in der Vergangenheit zu suchen sind und auch das aktive Aufsuchen von anregenden Bedingungen (Ziegler, 2005) eine bedeutende Rolle für die kognitive Fähigkeitsentwicklung spielt.

In den zuletzt genannten Studien wird also zwar der Einfluss der kognitiven Fähigkeiten als dominant herausgestellt, zeitgleich aber auf eine ebenso große Wirkung multipler weiterer Faktoren verwiesen. Nicht umsonst folgern daher Ferrer und McArdle (2004), dass der Zusammenhang zwischen kognitiven Fähigkeiten und akademischen Leistungen komplexer sei, als anhand der Investitionstheorie von Cattell erklärt werde. In ihrer Längsschnittstudie stand der Zusammenhang zwischen kognitiven Fähigkeiten und Leistungsfaktoren in der Kindheit bis zum Erwachsenenalter im Zentrum. Mit einer Stichprobengröße von 672 Kindern erfassten sie die quantitativen Fähigkeiten und allgemeines akademisches Wissen und untersuchten, inwiefern flüssige und kristallisierte Fähigkeiten spätere Vorhersagen ermöglichen. Ihre Ergebnisse deuten darauf hin, dass vor allem fluide Denkfähigkeiten verantwortlich für Veränderungen in den späteren Leistungen sind, während die Leistungsdaten aber spätere kognitive Fähigkeiten nicht vorhersagen konnten. Mit dieser Studie stellte sich für einige Forscher die Frage, inwieweit bis dahin zu Grunde gelegte Messinstrumente überhaupt in der Lage sind, die zu erfassenden Konstrukte und ihre Beziehung

zueinander präzise darzustellen. An dieser Stelle soll nochmals auf die im Rahmen der Invarianzfrage vorgestellte Studie von Watkins et al. (2007) eingegangen werden. Das Anliegen von Watkins et al. (2007) ging von dem Interesse aus, die Frage, inwieweit kognitive Fähigkeiten unmittelbar Einfluss auf schulischen Leistungsmaße besitzen, unter gleichzeitiger Berücksichtigung gegenseitiger Pfade von schulischen Leistungsmaßen auf spätere kognitive Leistungen zu beantworten. Watkins et al. (2007) versuchten dieser Herausforderung als Erste gerecht zu werden. Sie konzipierten eine Längsschnittstudie (2007) über zwei Messzeitpunkte und untersuchten die Pfade von Leistungsergebnissen zu Bildungsleistungen in gegenseitiger Richtung, um ein deutlich klareres Bild zur zeitlichen Präzedenz der Konstrukte zu geben. Sie ist eine der wenigen Publikationen, die die zeitliche gegenseitige Beeinflussung kognitiver Fähigkeiten und Schulleistungen thematisieren. Sie analysierten Daten von 289 Kindern mit sonderpädagogischen Fragestellungen, die hinsichtlich ihrer Förderlichkeit untersucht werden sollten. So führten sie in einem Abstand von 2,8 Jahren den Wechsler Intelligenztest für Kinder durch und weitere fünf Tests schulischer Leistungen. Die Schüler waren zum Testzeitpunkt 1 durchschnittlich neun Jahre alt und zum Testzeitpunkt 2, der zugleich der letzte Testzeitpunkt war, im Durchschnitt zwölf Jahre alt. Sie konnten mit der Analyse anhand eines Strukturgleichungsmodells belegen, dass hauptsächlich kognitive Fähigkeiten die späteren Schulleistungen bestimmen, während umgekehrt Leistungsergebnisse zukünftige Intelligenz Leistungen nicht wesentlich beeinflussen. Im Rahmen der Untersuchungen betonen Watkins et al. (2007) die zunehmende Bedeutung von Strukturgleichungsmodellen, die die Beziehungen zwischen Schulleistungen und kognitiven Fähigkeiten präziser als bisherige Analysemethoden bestimmen könnten und kritisiert die mangelnden empirischen Belege zu den zeitlichen Zusammenhängen zwischen kognitiven Fähigkeiten und Schulleistungen bei vorherigen Studien. Ihre eigenen Studienergebnisse werden jedoch durch die geringe Stichprobengröße von 289 eingeschränkt. Außerdem konnten zum zweiten Messzeitpunkt nicht mehr alle Kinder erneut getestet werden, da sie zum Teil die Sonderschule nicht mehr besuchten. Weiterhin handelte es sich um eine vorausgelesene Stichprobe von Kindern mit sonderpädagogischen Fragestellungen, die daher nur eingeschränkt repräsentativ ist, so Watkins et al. (2007) selbst. Die Studie umfasste ferner nur zwei Messzeitpunkte, wurde aber auf hohem methodischem Niveau mittels der Spezifikation von Strukturgleichungsmodellen sowie unter Prüfung der Invarianzbedingungen analysiert. Sie ist einer der ersten Versuche, die Beziehungen zwischen beiden Konstrukten präzise zu analysieren. Einschränkend jedoch muss gesagt werden, dass die Invarianzbedingungen nicht vollständig für die Bildungsleistungstests erfüllt waren. Bereits 2001 konnten

sie zeigen, dass die Testung der Invarianzbedingungen entscheidend für die Qualität der Aussagen sei, so Watkins, Lei und Canivez (2001). Die Invarianz der kognitiven Tests war gegeben, jedoch war die Invarianz der Leistungstests über die Zeit nicht eindeutig geklärt. So schlussfolgert Watkins, dass Intelligenz einen direkten Einfluss auf künftige Leistungsergebnisse besitzt, während Leistungsergebnisse zukünftige Intelligenzergebnisse nicht wesentlich beeinflussen. Damit stimmen sie abschließend mit der Position von Jensen (2000) überein. Sie interpretieren somit, dass kognitive Fähigkeiten und Leistungsmaße keine identischen Konstrukte sind.

2.2.4.1 Forschungsdebatten zu kreuzverzögerten Effekten kognitiver Fähigkeiten und schulischer Leistungen

In der Frage der gegenseitigen Beeinflussung, die Watkins et al. (2007) beschäftigte, blieben die Autoren letztlich die Antwort schuldig. Daher bleibt die hauptsächlich interessierende Frage, inwiefern sich schulische Leistungen auf die kognitiven Fähigkeiten zu einem späteren Zeitpunkt auswirken oder – aus der anderen Perspektive – inwiefern kognitive Leistung die späteren Schulleistungen beeinflussen ebenso unbeantwortet wie die Frage, inwieweit sich gegenseitige Effekte von früheren Leistungen auf spätere Zeitpunkte nachweisen lassen. Für das Grundschulalter insbesondere existiert noch keine Studie, die über zwei Messzeitpunkte hinausgeht und dabei messtheoretische Grundfragen, wie die Invarianzprüfung oder die Messfehlerbereinigung, berücksichtigt, sodass zuverlässige Aussagen bislang nicht vorliegen. Deshalb ist diese Arbeit der Versuch, den Faden von Watkins aufzunehmen. Einige Studien haben diesen Versuch im Ansatz bereits ebenfalls angedacht. So untersuchten Helmke und Weinert (1997), mittels einer im Rahmen der Scholastik-Studie durchgeführten Längsschnittstudie, inwiefern kognitive Determinanten die Lernfortschritte beeinflussen. Das Ziel war also die Rolle der allgemeinen Intelligenz als Lerndeterminante zu ermitteln (Weinert, 2001; auch Weinert & Helmke, 1997). Die Studie zu den Zusammenhangsmaßen zwischen mathematischen Schulleistungen und Intelligenz umfasste zu drei Messzeitpunkten den Zeitraum vom Kindergarten bis zum Ende der Grundschulzeit in der vierten Klassenstufe. Sie stellten fest, dass der Einfluss der Intelligenz für schulische Leistungserfolge im Laufe der Grundschulzeit mit zunehmender mathematischer Kompetenz abnehme. Dabei ist die Bedeutung der Intelligenz umso höher, wenn die Lerninhalte neu sind und wenig Vorwissen in dem Bereich verfügbar ist. Die Bedeutung des Vorwissens betonen sowohl Helmke und Weinert, als auch die Expertiseforscher. Sie konnten den Einfluss von früheren mathematischen Kompetenzen auf spätere mathematische Kompetenzen bestätigen, ohne

dass dieser jedoch allein für die Leistungsentwicklung von Schülern verantwortlich sei. Sie fanden einen zunehmend bedeutsamen Einfluss inhaltspezifischer Vorkenntnisse und einen abnehmenden Einfluss kognitiver Determinanten über die Zeit (Helmke & Baumert, 1997 in Köller & Baumert, 2002). Renkl (1996), Hasselhorn und Grube (1997) stimmen mit dieser Position überein und formulieren, dass die Rolle der Intelligenz deshalb abnehme, weil das erworbene Wissen über die Zeit den Erwerb neuer Lerninhalte begünstige. Diese Position wird auch teilweise seitens der Expertisetheorie gestützt. Nicht alle Expertiseforscher lehnen den Einfluss der kognitiven Steuerung als Wissensbasis angeeigneter Lernprozesse ab. Im Verlauf der Lernprozesse von Schülern, so die Interpretation von Weinert (2001), werden inhaltspezifische Vorkenntnisse bedeutsamer, da neue Informationen unter Rückgriff auf bereits Gelerntes zunehmend leichter vernetzt werden können. Schüler mit weniger Vorwissen können demnach weniger auf bereits Gelerntes zurückgreifen. In den von Weinert (2001) und auch Weinert und Helmke (1997) berichteten Ergebnissen der Scholastik Studie zum Einfluss des Vorwissens und kognitiver Determinanten von schulischen Leistungen offenbarte sich jedoch noch ein weiterer relevanter Befund. Es zeigen sich einerseits Einflüsse von kognitiven Fähigkeiten auf die Mathematikleistungen in Höhe von .30. Andererseits wurde ein, wenn auch geringer, wechselseitiger Einfluss von Mathematikleistungen auf die kognitiven Fähigkeiten sichtbar, obwohl dieser im Laufe der Zeit abnahm. Damit deuten diese Ergebnisse daraufhin, dass eher von einer wechselseitigen Beziehung zwischen Intelligenz und Mathematikleistungen auszugehen ist. In der Studie zur interindividuellen Entwicklung des mathematischen Kompetenzerwerbs über die Zeit belegen Helmke und Weinert (1997) den Einfluss früherer mathematischer Kompetenzen, die im Kindergartenalter erworben wurden, auf spätere mathematische Leistungen in der ersten Klassenstufe mit .45 und von der ersten Klassenstufe auf Mathematikleistungen zum Ende der vierten Klassenstufe mit .63. Auch Schneider et al. (2013) betonen die dominante Rolle der Schule beim Erwerb mathematischer Fähigkeiten ebenso wie Weinert (2001), und beide konstatieren, dass der Erwerb mathematischer Leistungsziele durch kumulatives, aufbauendes und vernetztes Lernen erreicht werde (Weinert, 2001). Dafür ist der Lernort Schule am besten geeignet, da diese Art Lernen durch professionelles Lehrerhandeln am ehesten erreicht werden kann. Für den förderlichen Einfluss der Schulbildung als Lernumgebung sprechen auch empirische Ergebnisse, die Zusammenhänge mit dem späteren beruflichen Status (Colom & Flores-Medosa, 2007; Mackintosh, 2011) und dem Einkommen (Berger & Kahlert, 2008; Weinert & Hany, 2000) berichten. Werden also schulische Einflüsse auf den mathematischen Kompetenzerwerb in den Mittelpunkt der Betrachtung gerückt, so können

diese kaum überschätzt werden. So berichteten auch Krajewski und Schneider (2009) aus einer Vierjahreslängsschnittstudie, dass für den Erwerb von mathematischen Kompetenzen auch mathematikspezifische Vorläuferfertigkeiten, wie sie im Vorschulalter erworben werden, eine Rolle spielen. Die mathematischen Leistungen wurden am Ende der ersten und vierten Klassenstufe erhoben. Die im Vorschulalter erworbenen Vorläuferfertigkeiten können bis zu 26 % der späteren Mathematikleistungen erklären. Die Autoren können aber keine dominante Rolle der Intelligenz dabei feststellen, sondern betonen Einflüsse wie den sozioökonomischen Status. Für das Kindergartenalter bis zum Ende der ersten Klassenstufe untersuchten Bossaert et al. (2011) diese Fragestellung. Sie stellten eine indirekte Wirkung kognitiver Fähigkeiten in Höhe von .18 auf die Eingangsleistungsmaße im Kindergartenalter bei den untersuchten 153 Kindern der Stichprobe fest. Der direkte Einfluss der Ausgangsleistungsmaße (Achievement) auf die Achievementmaße zum Ende der ersten Klassenstufe war mit einer Höhe von .46 am höchsten, die ferner durch den indirekten Einfluss kognitiver Fähigkeiten mit .18 moderiert wurden. Im Zuge dieser Studie erwiesen sich weiter insbesondere der Einfluss des sozioökonomischen Status⁴ und die Beteiligung im Unterricht als bedeutsame beteiligte Faktoren (Bossaert 2011). Neuere Forschungen unterstützen also die Annahme einer wechselseitigen Beziehung zwischen kognitiven Fähigkeiten und Leistungen, die wiederum Einfluss auf die kognitive Entwicklung nehmen (Rindermann, Flores-Medozza & Mansur-Alves (2010)). Die Autoren fanden in brasilianischen und deutschen Datensätzen sich gegenseitig verstärkende Einflüsse von kognitiven Fähigkeiten und Wissen. Sich gegenseitig verstärkende Effekte wurden auch von van der Maas et al. berichtet (2006). Sie gehen davon aus, dass schulisches Lernen kognitive Prozesse fördert, die wiederum Auswirkung auf die kognitive Entwicklung zeigen. Auch Messick (1984) nennt Belege für die Interaktion zwischen kognitiven Fähigkeiten und Wissensstrukturen während des Lernprozesses. Von dieser sich gegenseitig verstärkenden Natur bei der Entwicklung von akademischem Wissen berichtet auch Sternberg (1985). Er nimmt an, dass metakognitive Fähigkeiten und Lernfähigkeiten sowohl den kognitiven Fähigkeiten als auch den schulischen Leistungsmaßen zugrunde liegen. Für die Annahme gegenseitiger Einflüsse sprechen auch die Ergebnisse von Soares et al. (2014). In ihrer Längsschnittstudie stellten sie die Vorhersagbarkeit von schulischen Leistungen am Ende der Klassenstufe neun durch die kognitiven Fähigkeiten und die schulische Leistung zu Beginn der Untersuchung in Klassenstufe sieben in den Vordergrund ihrer Fragestellung. Mittelschüler der siebten bis neunten Klassenstufe aus Portugal absolvierten über drei Jahre drei kognitive Fähigkeitstests, in

denen das abstrakte, das numerische und das verbale Niveau erhoben wurde. Zugleich wurden Schulleistungen in Form von Schulnoten erfasst. Soares et.al. (2015) konnten nachweisen, dass die Intelligenz und die Schulnoten als zwei unabhängige Prädiktoren der Leistungen extrahiert werden konnten. Dabei ließ sich die abschließende akademische Leistung am besten durch die Intelligenz vorhersagen, wobei die Schulnoten, wenn sie als Vermittlungseffekt mit einbezogen werden, einen erheblichen Anstieg der prädiktiven Gültigkeit der Intelligenz erlauben. Die indirekte Wirkung der anfänglichen schulischen Leistungen erklärte etwa 71 % der Varianz der endgültigen schulischen Leistungen. Doch auch die direkte Wirkung der Intelligenz auf die anfängliche schulische Leistung ist signifikant und erklärt 50 % der anfänglichen Schulleistungen (Noten zu Beginn der Klassenstufe sieben). Wiederum war die direkte Wirkung von Intelligenz auf die Schulleistungen am Ende der Klassenstufe neun nicht signifikant. Unter Hinzunahme der anfänglichen schulischen Leistungen erfährt die prädiktive Geltung der Intelligenz einen erheblichen Zuwachs. Die Ergebnisse von Soares lassen also die Annahme zu, dass sowohl kognitive Fähigkeiten als auch schulische Leistungen autonome Variablen sind, die jedoch beide gemeinsam letztlich die Schulleistungen vorherzusagen vermögen. Soares erklärt die Vermittlungsrolle von bisherigen schulischen Leistungen und Intelligenz mit der Investitionstheorie nach (Cattell, 1971). Demnach spielt die fluide Denkfähigkeit eine relevante Rolle für das Lernen und hilft den Schülern, neue Fähigkeiten und Fertigkeiten zu erwerben, die erforderlich sind, um sich neue schulische Wissens Elemente anzueignen (Cattell, 1971; Primi et al., 2010). Insofern wird die fluide Intelligenz als ein Faktor des Lernens betrachtet, der sich vor allem in neuartigen Situationen bemerkbar macht und die Aneignung von Fähigkeiten und Kenntnissen erleichtert. Die erworbene kristallisierte Intelligenz, die im Zusammenhang mit dem erworbenen Bildungswissen steht, wirkt schließlich auf die fluiden Intelligenzanteile zurück und beeinflusst diese wiederum positiv. So erleichtert die kristallisierte Intelligenz auch die Erfassung komplexer Probleme. In der Intelligenzforschung machten bereits Cattell und Kline (1977) auf die Bedeutung der wissensbasierten (kristallisierten) Anteile von Intelligenz aufmerksam. Allerdings wurden diese Daten in einer Gruppe von Mittelstufenschülern erhoben. Für das Grundschulalter können daher keine Aussagen abgeleitet werden. Darüber hinaus waren die Invarianzbedingungen über die Zeit nur für die kognitiven Fähigkeitstests gegeben, nicht aber für die schulische Leistungsfähigkeit. Präzise Aussagen gelten daher nur für das angegebene Alter zwischen der siebten und neunten Klassenstufe. Weiterhin muss einschränkend gesagt werden, dass die Stichprobengröße recht übersichtlich war. Auch andere Forscher schlagen unterdessen vor, dass die Beziehung zwischen kognitiven

Fähigkeiten und schulischen Leistungen gegenseitiger Natur sei (Brody, 1997). Brody (1997) schlägt aber eine etwas andere Interpretation vor. Er geht davon aus, dass die enge Beziehung zwischen Intelligenztests und Bildungsniveau zwar auch durch sich gegenseitig beeinflussende Faktoren zustande kommt. Brody (1997) betont aber, dass Einzelpersonen, die eine akademische Laufbahn wählten, selbst wenn sie im Vergleich mit anderen Personen im Alter von 13 Jahren einen vergleichbaren Intelligenzquotienten besaßen, ihren Zugewinn an Intelligenz dadurch erreicht haben könnten, dass sie aus Interesse mehr Bücher lasen als andere Personen. Damit deutet er wiederum die Abhängigkeit der kognitiven Fähigkeitsentwicklung und des Bildungsstandes von Personeneigenschaften an. Er ist der Ansicht (1993), dass die Wahl der akademischen Schullaufbahn Einfluss auf die Entwicklung kognitiver Fähigkeiten des Einzelnen haben könnte (1993). Wiederum gehen Merz, Renner und Ehlers, (1985) davon aus, dass auch die kognitiven Fähigkeiten selbst im Sinne kristallisierter und fluider Intelligenz bildungsabhängig seien. Diese Position vertreten auch Schneider und Stanek (2004). Stanovich (1986) machte sie als den in Kapitel 2 beschriebenen „Matthäus-Effekt“ bekannt, als er versuchte die Mechanismen aufzudecken, die dazu führen, dass fähigere Kinder immer zügiger immer fähiger werden, während Kinder mit Lernproblemen weiterhin Lernschwierigkeiten zu bewältigen haben, bzw. sogar zunehmend schlechter im Vergleich zu Gleichaltrigen werden. Stanovich untersuchte diesen Effekt anhand der Entwicklung des Wortschatzes, der das Leseverständnis erleichtert. Das Lesen wiederum fördert die Erweiterung des Wortschatzes. Diese wechselseitige Beziehung kann große Leistungsunterschiede innerhalb eines Klassenzimmers erklären. Hinzu kam, dass weniger motivierte Kinder im Durchschnitt auch schlechter lesen konnten. Die Unterschiede zeigten sich im Wortschatz bei dem Neuerwerb von Wörtern. Die Kinder, die wiederum neue Wortbedeutungen schneller erwerben, können demzufolge auch mit noch unbekannten Wörtern besser umgehen und sind motivierter beim Lesen. Im Gegensatz dazu haben weniger schnelle Leser eine langsamere Entwicklung des Wortschatzes, welche sich wiederum nachteilig auf den weiteren Erwerb der Lesefähigkeit auswirkt. Demnach sei festzustellen, dass die zukünftige Entwicklung von dem in der Vergangenheit erworbenen Wissen und von den kognitiven Grundvoraussetzungen abhängt. Je höher das Niveau vormaliger kognitiver Fähigkeiten, um so eher lasse sich auch neu erworbenes Wissen effizient mit bereits bestehenden Wissensdomänen verknüpfen, und in später benötigten relevanten Situationen zügig zur Verfügung stellen. Die Effizienz dabei bewirke wiederum freiwerdende kognitive Ressourcen, die im interindividuellen Vergleich zügiger und effizienter für weitere kognitive Anforderungen genutzt werden können. Die bisherige Forschung weist darauf hin, dass vor

allem kognitive Fähigkeiten ursächlich sein können, aber es werden auch Umweltfaktoren und sozioökonomische Faktoren diskutiert. Stanovich (1986) betont die Bedeutung des Verständnisses der Ursachen, um sozialpolitische Intervention einzuleiten. So empfiehlt er, die unterschiedlichen Verarbeitungsmuster zu beachten, die zur Bewältigung von Lernschwierigkeiten eingesetzt werden können. Die Autoren Schneider und Stefanek (2004), die den Matthäus-Effekt als „Schereneffekt“ bezeichnen, untersuchten, ob sich die oben beschriebenen zunehmenden individuellen Unterschiede über einen Zeitraum von 13 Jahren nachweisen lassen. Kinder im Alter von vier Jahren führten einen umfassenden Intelligenztest durch. Zudem wurden die Kompetenzen im Lesen, Rechtschreiben und in Mathematik erfasst. Im Alter von 17 Jahren ließen sich jedoch nur Schereneffekte für die verbale Intelligenz, das formale Denken sowie die Rechtschreibleistung nachweisen. Eingangs bestehende bedeutende Differenzen änderten sich im Laufe der Zeit kaum. Ob sich dieser Effekt auch bei Kindern mit vorzeitiger Lesefähigkeit zeigte, untersuchten Cain und Oakhill (2011). Sie berichten Ergebnisse einer Längsschnittstudie zur Leseentwicklung von Kindern, die im Alter von 8, 11, 14 und 16 Jahren erfasst wurden. Die Kinder wurden im Alter von acht Jahren unterschiedlichen Gruppen von Lesern mit gutem bzw. schlechtem Leseverstehen zugeordnet. Die schlechteren Leser hatten einen geringeren Vokabelwortschatz im Vergleich zu den guten Lesern, aber nicht in der Wort-Lese- oder Leseverständnisfähigkeit. Auch durch die Autoren Bast und Reitsma (1998) ließ sich der Matthäus-Effekt in Form reziproker Beziehungen hinsichtlich der Worterkennungsfähigkeiten feststellen. Einige Forschungsergebnisse der aufgeführten Studien zeigen außerdem, dass Kinder einen durchschnittlichen Intelligenzquotienten in der Ausgangslage besaßen, dessen weiterer Verlauf bei intermittierender Schulbildung kumulative Effekte zeigt und damit in Richtung des beschriebenen Matthäus-Effekts deutet.

Die dokumentierten Zusammenhänge zwischen kognitiven Fähigkeiten und schulischen Leistungen werden in jüngster Forschung also auch als sich gegenseitig beeinflussende Faktoren betrachtet (Perleth, 1997; Watkins, 2007; Soares, 2015). Obwohl zum Beispiel Watkins den wechselseitigen Einfluss nicht endgültig empirisch bestätigen konnte, gibt es hinreichend theoretische Begründungen, die wechselseitige Beeinflussung anzunehmen. So erweitere der Wissenserwerb in der Schule beispielsweise die Problemlösefähigkeiten von Personen, indem sie zum Beispiel wissensbasierte Denkstrukturen effizienter zu nutzen lernen (Ceci, 1991). Im Zuge dessen können sie in der Schule vermittelte Wissens-elemente erfolgreich auf andere Bereiche transferieren. Dadurch wird das kognitive Niveau

der Personen erneut erweitert. So sei der Unterricht auf die Vermittlung entsprechender Methoden fokussiert, deren Ziel die Erweiterung dieser Problemlösefähigkeiten ist. Nach der Theorie, die eine gegenseitige Beeinflussung annimmt, können Schüler die vermittelten Lerninhalte im Gedächtnis behalten, relevante Informationen abrufen und Prozessinformationen transferieren. Auf der anderen Seite gelingt es den Schülern, kognitive Ressourcen für den Erwerb neuer Wissensstrukturen zu schaffen, die wiederum den weiteren Wissenserwerb erleichtern. Schülern mit hoher Ausprägung kognitiver Fähigkeiten fällt es dabei vergleichsweise leichter, kognitive Ressourcen für den Erwerb neuer Wissensstrukturen zu schaffen. Das wiederum bewirkt einen zusätzlichen Wissensvorteil. Diese Schüler können die vermittelten Lerninhalte effektiver im Gedächtnis behalten, der Abruf relevanter Informationen ist effizienter und auch der Transfer von Prozessinformationen gelingt zumeist zügiger. Dieser Vorgang des Lernens wird als kumulatives Lernen (Grube, 2008) bezeichnet. Die Grundannahme dieser Position ist, dass vor allem akademische Leistung im Sinne schulischer Bildung die spätere akademische bzw. schulische Leistung bestimmt (Soares, 2015). Schulleistungen sind also von früheren Leistungen abhängig und damit auch von früher investierter Förderung im Sinne von erworbenem Wissen bzw. Vorwissen (Helmke & Weinert, 1997). Die Vertreter dieser Position gehen somit davon aus, dass das auf die eine oder andere Weise akkumulierte Vorwissen einen gravierenden Einfluss auf die Leistungen im schulischen Bereich ausübt und zum anderen auch die kognitiven Fähigkeiten beeinflusst (Ericsson, 1996; Ericsson, Krampe & Tesch-Römer, 1993). Allerdings befanden sich bislang die Einflüsse schulischer Vorwissenskomponenten nur in wenigen Untersuchungen explizit im Blickfeld (Perleth, 1997). Perleth (1997) hingegen bildet hier eine Ausnahme. In seinem im Theorie-Teil erläuterten Münchner dynamischen Begabungs-Leistungs-Modell (1997) werden Wissenserwerbsprozesse explizit berücksichtigt, denn er stellt die theoretischen Überlegungen interagierender Effekte zwischen Vorwissen, schulischen Leistungen und kognitiven Fähigkeiten in den Vordergrund.

2.2.4.2 Kreuzverzögerte Effekte – Zusammenfassung, Ableitung und Hypothesen

Fasst man den Forschungsstand nun in Bezug auf die Frage nach den Beziehungen zwischen Intelligenz und akademischem und schulischem Erfolg zusammen, lässt sich feststellen, dass trotz zahlreicher Studien zwar die grundlegende Übereinstimmung darüber auszumachen ist, dass es zu einer Beeinflussung kommt. Auch die Existenz einflussnehmender Faktoren ist nicht grundsätzlich strittig. Doch die Richtung und Stärke des Einflusses ist nach

wie vor nicht eindeutig geklärt. Vertreter einer Position gehen von einer vorwiegend unabhängigen Wirkung auf spätere Leistungen des gleichen Konstrukts aus. Schulleistungen bedingen demnach Schulleistungen und seien relativ unabhängig von kognitiven Leistungen (Geary, 1995; Ericsson, & Smith, 1991; Ericsson, Krampe & Tesch-Römer, 1993). Vertreter einer anderen Position wie Brinch und Galloway (2011) oder Jencks (1973) führen jedoch weitere Hinweise an, dass durch formative Bildung selbst eine Steigerung der kognitiven Fähigkeiten von Kindern zu erzielen sei.

Einige Forscher vertreten die Annahme der multifaktoriellen Beeinflussung beider Konstrukte. Seitens der psychologischen Forschung wird auch dem Anregungsgehalt der familiären und schulischen Lernumwelt, sowie den nicht-kognitiven Persönlichkeitsmerkmalen (z. B. Perleth, 1997) eine wesentliche Bedeutung beigemessen. Moderatoren der Lernumgebung, wie Instruktionsqualität des Unterrichts oder kritische Lebensereignisse, sind ebenfalls bedeutsam (Perleth, 1997). Schulleistungen sind nach den oben genannten Ergebnissen zu einem gewissen Teil von früher investierter Förderung im Sinne von erworbenem Wissen abhängig (Perleth, 1997). Wenngleich der Einfluss der Motivation und der Persönlichkeitsmerkmale auf die schulischen (Mathematik)leistungen belegt ist, stellt die psychologische Forschung jedoch die Bedeutung der kognitiven Fähigkeiten für den Leistungskontext heraus und sieht vornehmlich diese kognitiven Voraussetzungen als Disposition zur Leistungserbringung (Stern (1997, Helmke, 1997, 2008; Heller, 1991). Der erfolgreiche Abruf früherer (Wissens-)Informationen zum Einsatz aktueller Anforderungen ist wiederum von Prozessen der Informationsverarbeitung (Jensen, 1998) bzw. von der Fluidität der Denkfähigkeit im Sinne von Cattell (1971) und damit von kognitiven Fähigkeiten abhängig. Versucht man die Fäden zur Beantwortung der Eingangsfrage zusammenzuführen, bleibt festzuhalten, dass der bisherigen Forschungslage nicht eindeutig zu entnehmen ist, ob schulische Leistungen kognitive Fähigkeiten beeinflussen, kognitive Fähigkeiten schulische Leistungen bestimmen oder ob und inwiefern die Effekte gegebenenfalls eine gegenseitige Wirkung entfalten. Lässt man die Studien Revue passieren, kann dieses Resultat nicht überraschen. Schon Ceci (2001) weist auf fundamentale methodische Probleme der vorliegenden Studien hin. Er wendet ein, dass die Ergebnisse häufig auf empirisch zu grob erhobenen Daten basieren, um eine detailliertere Analyse zu erlauben. Nach seiner Metaanalyse von zweihundert Studien, die sich den Zusammenhängen zwischen kognitiven Fähigkeiten und mathematischen Schulleistungen widmeten, betont er deren begrenzte Aussagekraft (Ceci, 1991). Zunächst erwähnt er Probleme bei der Auswahl der Stichproben und bei der Konzeption entsprechender Studiendesigns. Auch betont er Messprobleme, denn, so Ceci

(1991), die Auswirkung von Schulbildung auf Intelligenz könne deshalb zustande kommen, weil das Muster der Beziehungen zwischen schulischen Variablen und Testergebnissen ähnlich wie das Muster der Beziehungen zwischen schulischen Variablen und Intelligenzscores sei. Dies deute darauf hin, dass Intelligenztestleistungen, ebenso wie die Testergebnisse schulischer Bildung, durch akademische Leistung beeinflusst sind, wie z. B. den Erwerb entsprechender Kompetenzen im Umgang mit den Testverfahren selbst.

Er stellte außerdem fest, dass die meisten Studien auf korrelativer Ebene analysiert worden sind, was letztlich nicht zur Beantwortung der interessierenden Frage führen könne, da Kausalbeziehungen nicht aufgedeckt werden und zweideutige Schlussfolgerungen (Ceci, 1991) die Konsequenz sind. Obgleich also die hohen Korrelationen von .77 (Ceci, 1991) zwischen kognitiven Fähigkeiten und schulischen Leistungen gut dokumentiert sind, ist die Wirkrichtung zwischen Schulbildung und kognitiven Fähigkeiten trotzdem aus diesen Gründen nur schwer bestimmbar.

Noch so hohe Korrelationen können nicht beantworten, welches Konstrukt als ursächlich für den Zusammenhang angesehen werden kann. Diese Kritik trifft auch die Studie von Deary et al. (2007), die von sehr hohen Korrelationen berichten. Ihre Ergebnisse basieren auf einer 5-Jahres-Längsschnittstudie mit 70000 Kindern. Die Korrelation zwischen der allgemeinen Intelligenz im Sinne von Spearmans g im Alter von 11 Jahren und den Bildungsleistungen in verschiedenen Schulfächern im Alter von 16 Jahren hatte eine Höhe von .81. Dabei war der Einfluss auf die Mathematikleistungen mit einem Regressionskoeffizienten von .58 bis zu .77 am höchsten. Bemerkenswert ist aber, dass ein sehr großer Anteil der Varianz eben nicht auf allgemeine Intelligenzfaktoren zurückzuführen war. Als Erklärung benennen die Autoren zum einen Limitationen durch Messfehler, aber eben auch systematische Faktoren, die nicht berücksichtigt wurden. Dazu gehören Persönlichkeitsfaktoren wie Engagement, Motivation und Anstrengungsbereitschaft oder auch das Ausmaß der elterlichen Unterstützung und der Anregungsgehalt der Lernumwelt, die Qualität der Lehre und schulisches Ethos. Sogar Dearys Analysen der sehr hohen Korrelationsbeziehungen von bis zu .81 konnten zeigen, dass etwa 50 % – 60 % der Varianz der Schulleistungen durch kognitive Faktoren (g -Faktor) statistisch erklärt werden können, während ein ebenso hoher Anteil der Varianz nicht auf eine allgemeine kognitive Fähigkeit g entfiel (Deary, 2007). Ein Teil dieser verbleibenden Varianz könne, so Deary selbst, durch Messfehler erklärt werden. Die wenigen Längsschnittuntersuchungen, mit denen versucht wurde, die Wirkrichtung und die Stärke des Einflusses kognitiver Fähigkeiten auf zukünftig Schulleistungen deutlich zu machen, sind folglich nicht uneingeschränkt belastbar.

Von statistischen Problemen der vorliegenden Studien berichten Winship und Korenman (1997). Sie reanalysierten die Herrnstein und Murray Daten aus dem Jahr (1994) und hatten Schwierigkeiten mit der Reanalyse, da diese nicht das Alter der Personen zum Zeitpunkt des ersten Messzeitpunktes berücksichtigte. Des Weiteren gab es eine ungewöhnliche Handhabung im Umgang mit fehlenden Daten.

Bisher berichtete Ergebnisse basieren häufig zudem auf vorausgelesenen Stichproben bzw. betrachten weitaus längere Zeiträume, sodass die Ergebnisse auch durch zahlreiche andere Faktoren moderiert werden und eindeutige Rückschlüsse empirisch nicht mehr gelingen. Shanley (2016) betont daher, dass eine präzise Messung und Modellierung von schulischen Leistungen entscheidend sei, um messfehlerfreie Ergebnisse zu ermitteln und moniert, dass oft vorausgelesene und/oder zu geringe Stichproben in die Analysen eingehen.

Datenanalysen aufgrund kleiner Stichproben und Nichtberücksichtigung verschachtelter Datenstrukturen verursachen demnach weitere methodische Probleme. Für präzise Aussagen zu den Zusammenhängen und insbesondere zu den tatsächlichen Wirkrichtungen ist also die Qualität der Stichprobe entscheidend, die möglichst unvorausgelesen und umfangreich genug sein sollte. Die Stichprobengröße vorliegender Studien ist aber häufig nicht ausreichend, zudem viele von ihnen vorausgelesen sind.

Auch die gemischte Analyse von Daten, die mit unterschiedlich konzeptionierten Testverfahren erhoben wurden, ist problematisch. Wenn beispielsweise einige Testverfahren allgemeine Intelligenzfaktoren erheben, während andere bereichsspezifische Fähigkeiten erfassen, wird die messfehlerfreie Interpretation erschwert. Große Zeitabstände zwischen den Analysezeitpunkten bergen zusätzlich die Gefahr unberücksichtigter Drittvariableneinflüsse.

Auch werden meist kognitive Tests auf globaler und undifferenzierter Ebene erfasst, in der beispielsweise die hierarchische Organisation kognitiver Fähigkeiten, wie nach dem Modell von Carroll (1993), außer Acht gelassen wurde. Es ist außerdem schwierig, die Ursachen aktueller Beobachtungen in der Vergangenheit zu suchen, da möglicherweise das aktive Aufsuchen von anregenden Bedingungen (Ziegler, 2005) eine bedeutende Rolle auf die spätere kognitive Fähigkeitsentwicklung spielt. Die diskutierten Forschungspositionen entstehen zum Teil aufgrund der Retrospektive, aus der die Daten analysiert werden. Dabei ist der Einfluss von Messfehlern häufig enorm. Aussagen zum Einfluss früherer Konstrukte auf spätere Konstrukte sind eben nur dann möglich, wenn mehrere Messzeitpunkte erhoben werden, die untereinander in Beziehung gesetzt werden können (und nicht nur Korrelationen zwischen zwei als Querschnitt angelegten Studien ermitteln). Insbesondere lassen also

vor allem die Studien, die ihre Ergebnisse aufgrund von querschnittlichen Studiendesigns gewannen, methodische Fragen offen. Auch zwei zu zwei Messzeitpunkten erhobene kognitive und schulleistungsrelevante Parameter sind noch keine Längsschnittstudie im eigentlichen Sinne, wenn sie keine pfadanalytische Beziehung zwischen vorherigen und späteren Daten zulassen. Bislang durchgeführte Längsschnittuntersuchungen aber umfassen nur zwei Messzeitpunkte (Widaman, et al., 2010; Watkins et al., 2007) und können daher keine Aussage zu dem zeitlichen Verlauf kreuzverzögerter Effekte machen. Längsschnittstudien für den Teilbereich mathematische Schulleistungen und kognitive Fähigkeiten bleiben die Ausnahme, die Regel sind Studien, die auf manifester Ebene korrelativ argumentieren. Studien, die explizit das Grundschulalter in den Fokus ihrer Forschungsbemühungen stellen, haben ohnehin Seltenheitswert. Allgemein wird der Aussagegehalt existierender Studien aufgrund vorausgelesener, kleiner Stichproben, unbeachteter Messfehlereinflüsse, retrospektiver Analysen und zu langer Zeiträume zwischen den Erhebungen gemindert. Es bestehen folglich in diesem Bereich methodische Forschungslücken, die für die inhaltlichen Forschungslücken mitverantwortlich zu machen sind. Es fehlen bislang Forschungsergebnisse, die den gegenseitigen Verlauf von Schulleistungsmerkmalen und kognitiven Fähigkeiten unter gleichzeitiger Berücksichtigung der gegenteiligen Effekte von früheren Mathematikleistungen auf kognitive Fähigkeiten und in der umgekehrten Einflussrichtung integrieren.

An dieser Stelle sei auf das Problem der Kausalität von Aussagen hingewiesen. Kausale Aussagen sind zumeist Ziel von Längsschnittuntersuchungen. Doch um kausale Aussagen treffen zu können sind nach Duncan, Duncan und Strycker verschiedene Bedingungen unabdingbar (Duncan et al. 2006; in Eid, Gollwitzer & Schmitt, 2013). Zum einen müssen Experimente durchgeführt werden, denen wiederholte Messungen einer Person über die Zeit zugrundeliegen. Diese erfordern eine enge zeitliche Distanz, um Varianzveränderungen tatsächlich durch Unterschiede im Entwicklungsverlauf feststellen zu können. Dabei ist die Zeiteinteilung zwischen den Messzeitpunkten entscheidend. Zusätzlich müssen den Längsschnittdaten die angenommene Ursache und der Effekt miteinander verbunden sein, die Ursache muss dem Effekt zeitlich vorausgehen und andere Erklärungen müssen ausgeschlossen werden können (Duncan et al. 2006; in Eid, Gollwitzer & Schmitt, 2013). Die zugrundeliegende Datenstruktur erlaubt zwar die Erfassung situationsspezifischer Schwankungen, die Verbindung von Ursache und Effekt über vier Messzeitpunkte durch das längsschnittliche Design, während andere Erklärungen nicht vollständig ausgeschlossen werden können, da der zeitliche Abstand nicht die Bedingungen eines Experiments erfüllt, sodass letztlich

kausale Schlussfolgerungen nicht möglich sind. Dennoch sind weitaus präzisere Aussagen über gegenseitige Einflüsse der untersuchten Konstrukte möglich, wenn die Analysen auf latenter Ebene erfolgen, denen möglichst viele Messzeitpunkte zugrundeliegen.

Untersuchungen zur gegenseitigen Beeinflussung interagierender Moderatoren betrachteten bislang nur zwei Messzeitpunkte (Watkins et.al, 2007). Dabei ist die differenzierte Betrachtung eines längeren Zeitraums interessant, um zur Leistungsentwicklung von Kindern im Grundschulalter differenziertere Aussagen machen zu können.

Die 4. Fragestellung widmet sich der gegenseitigen Beeinflussung von kognitiven und mathematischen Fähigkeiten, denn bislang wurde die Möglichkeit der gegenseitigen Beeinflussung nur innerhalb von zwei Messzeitpunkten untersucht. Folglich können für einen Zeitraum, der die gegenseitige Beeinflussung kognitiver Fähigkeiten und Mathematikleistungen über vier Messzeitpunkte betrachtet, noch keine Aussagen über den zeitlichen Verlauf kreuzverzögerter Effekte getroffen werden. Daher wird die vierte Fragestellung wie folgt formuliert:

Die Fragestellung 4 soll klären, ob kognitive Fähigkeiten aus früheren Messungen die mathematischen Schulleistungen zu späteren Zeitpunkten beeinflussen und ob mathematische Leistungen aus vorherigen Messzeitpunkten auf kognitive Leistungen späterer Zeitpunkte wirken. Es soll festgestellt werden, ob von einer gegenseitigen Beziehung auszugehen ist und wie stark die Einflüsse kognitiver Fähigkeiten auf spätere mathematische Schulleistungen sind bzw. in welcher Stärke wiederum mathematische Schulleistungen auf kognitive Fähigkeiten späterer Zeitpunkte wirken.

Aus dieser Fragestellung werden folgende Hypothesen abgeleitet:

Hypothese 4a:

Es wird erwartet, dass kognitive Fähigkeiten die mathematischen Schulleistungen über die vier betrachteten Messzeitpunkte zu späteren Zeitpunkten beeinflussen.

Hypothese 4b:

Es wird von einer Wirkung früherer mathematischer Schulleistungen auf spätere kognitive Leistungen über die vier betrachteten Messzeitpunkte ausgegangen.

Hypothese 4c:

Es wird davon ausgegangen, dass sich kognitive Leistungen und mathematische Schulleistungen über die vier betrachteten Messzeitpunkte gegenseitig beeinflussen.

Hypothese 4d:

Es wird zudem erwartet, dass sich diese Effekte bereits von der ersten Klassenstufe an mit Wirkung auf die zweite Klassenstufe zeigen.

Hypothese 4e:

Es wird angenommen, dass sich die Effekte eines Konstrukts auf das jeweils andere im Laufe der Grundschulzeit verstärken, sodass erwartet wird, dass sich die größten gegenseitigen Effekte von der dritten zur vierten Klassenstufe des Gegenkonstrukts zeigen.

Ergebnisse dieser Analysen vermögen mit der Beantwortung dieser Frage einen Beitrag zur Klärung kreuzverzögerter Effekte von kognitiven Merkmalen und mathematischen Schulleistungen über den relevanten Grundschulzeitraum von vier Jahren mit vier Messzeitpunkten zu leisten. Dabei werden potentielle methodische Probleme mittels Spezifikation von Strukturgleichungsmodellen berücksichtigt, um die Richtung der Beeinflussung empirisch zu klären.

3 Empirischer Teil

In dieser grundlagenwissenschaftlichen Arbeit, die einen anwendungsrelevanten Beitrag leisten will, soll nun im nächsten Kapitel der Versuch unternommen werden, eine theoretische Klärung der oben abgeleiteten, noch unbeantworteten Fragen herbeizuführen und gleichzeitig auch methodische Forschungslücken, die oben dargelegt wurden, zu schließen. Es soll zusammenfassend geklärt werden, inwieweit das Konstrukt der Intelligenz unter Berücksichtigung von Messfehlern und auf der Ebene von latenten Variablen aus früheren Messungen vorhersagbar ist und zweitens, inwiefern sich spezifische Schulleistungen, hier mathematische Schulleistungen, aus früher gemessenen kognitiven Fähigkeiten ableiten lassen bzw. inwieweit sich möglicherweise kognitive Fähigkeiten aus mathematischen Fähigkeiten früherer Messzeitpunkte vorhersagen lassen. Dazu müssen, kurz rekapituliert, nun folgende Fragen empirisch adressiert werden:

- 1) Ist die Messung kognitiver Fähigkeiten und Mathematikleistungen vergleichbar über die Zeit?
- 2) Wie stabil sind die erfassten Konstrukte über die Zeit?
- 3) Sind kognitive Fähigkeiten alleine prädiktiv für Mathematikleistungen, oder kann von einer gegenseitigen Beeinflussung ausgegangen werden?
- 4) Ist die Erfassung kognitiver Fähigkeiten und Mathematikleistungen innerhalb der Messzeitpunkte partiell unabhängig?

Um aber der Aufgabe der Arbeit gerecht zu werden und tatsächlich die Wirkrichtung und die Stärke des Einflusses kognitiver Fähigkeiten auf zukünftige Schulleistungen deutlich zu machen, sind Langzeitstudien besonders geeignet. Für präzise Aussagen zu den Zusammenhängen und insbesondere zu den tatsächlichen Wirkrichtungen ist zudem die Qualität der Stichprobe entscheidend, die möglichst unvoreingenommen und umfangreich genug sein sollte. Darüber hinaus sind Aussagen zum Einfluss früherer Konstrukte auf spätere Konstrukte nur dann möglich, wenn mehrere Messzeitpunkte erhoben werden, die untereinander in Beziehung gesetzt werden können (und nicht nur Korrelationen zwischen zwei als Querschnitt angelegten Studien ermitteln). Zudem eignen sich latente Strukturgleichungsmodelle besonders für präzise Analysen, denn mit ihnen lässt sich der Einfluss von Messfehlern ermitteln. Ein besonderer Vorteil ist die komplexe Berücksichtigungsfähigkeit verschiedener Konstrukte auf Item- bzw. auf Subtestebene, da eben nicht nur allgemein „kognitive Fähigkeiten“ oder „Schulleistungen im Allgemeinen“ erfasst werden, sondern auch auf differenzierter Ebene unterschiedlich starke Einflüsse verschiedener Fähigkeitsbereiche des jeweiligen Konstrukts erfasst werden.

Diese Arbeit beabsichtigt daher, die oben genannten offenen theoretischen Fragen auf empirischer Ebene zu prüfen und methodische Unzulänglichkeiten zu vermeiden. Der empirische Teil ist als Längsschnitt mit vier Messzeitpunkten angelegt. Den Messungen liegen zwei theoretisch präzise standardisierte Messverfahren zugrunde. Sowohl das Konstrukt mathematische Schulleistungen als auch das Konstrukt kognitive Fähigkeiten wird auf bereichsspezifischer Merkmalsebene auf mehreren Subtestebenen erfasst, sodass eine hinreichende Unterscheidung von Personenfähigkeiten möglich ist. Die Fähigkeiten verschwinden so nicht in einem allgemeinen g der allgemeinen „Mathematikleistungen“, sondern es werden hierarchische Organisationen berücksichtigt. Den Messungen liegt eine hinreichend große Anzahl an Items zugrunde, die eine sehr genaue Analyse ermöglicht.

Es handelt sich um eine große unausgelesene Stichprobe und ein Analysedesign, das zuvor auf latenter Ebene die Invarianz berücksichtigte. Der Umgang mit fehlenden Daten

wird berichtet. Es werden Testverfahren eingesetzt, die jeweils auf demselben theoretischen Konzept beruhen und so die oben beschriebenen Stichprobenfehler in der Datenanalyse vermeiden. Auch die Kontrolle der Mehrebenenstruktur wurde berücksichtigt, da eine Verschachtelung der Daten durch die Clusterung in Schulen vorliegt. Zudem wurden konstante jährliche Untersuchungen vorgenommen, was die Gefahr unberücksichtigter Drittvariableneinflüsse mindert, aber dennoch genügend Zeit für die Erfassung der Feststellung von interindividuellen Unterschieden über die Zeit erlaubt. Es kann aufgrund der vier Messzeitpunkte sowohl eine Rückschau erfolgen, aber auch vom zweiten und dritten Messzeitpunkt aus gesehen eine prospektive Sicht eingenommen werden. Basis sind Strukturgleichungsmodelle. Der Vorteil dieser latenten strukturgleichungsmodellbasierten Faktormethoden ist, dass sie die Aufdeckung der latenten Beziehungen zwischen kognitiven und Leistungsvariablen und somit Aussagen über sonst unaufgeklärte Varianzen erlauben. Damit ist eine messfehlerbereinigte Schätzung möglich (Watkins, Lei & Canivez, 2007). Latente Strukturgleichungsmodelle sind im Vergleich zu anderen Modellen geeignet, differenzierte Beziehungen darzustellen. In dieser Arbeit werden konkret der KFT (bereichsspezifisch) und der DEMAT (stoffgebietsspezifisch) verwendet.

Die vorliegende Studie ist also die erste, die auf latenter Ebene und unter Berücksichtigung von zeitrelevanten Aspekten Zusammenhänge und Effekte zwischen mathematischen Schulleistung und kognitiven Fähigkeiten, die in standardisierter Form (!) erhoben wurden, im Längsschnitt analysiert, um die Identität der Konstrukte zu prüfen. Dabei erfolgt eine Konzentration auf die in der theoretischen Debatte am häufigsten kontrovers diskutierten Einzelkomponenten. Hierbei handelt es sich zum einen um kognitive Fähigkeiten, die häufig mittels Intelligenztests gemessen werden und so den weiteren Bildungsweg von Schülern determinieren. Zum anderen stehen die mathematischen Schulleistungen im Mittelpunkt, weil diese, ebenso wie Intelligenztests, am ehesten als Prädiktor für schulischen Erfolg betrachtet werden. Darüber hinaus wurden mathematische Schulleistungen in die Analysen der vorliegenden Arbeit zum einen deshalb aufgenommen, da sie als (vor)-wissensbasiert gelten (Geary, 1995, 2000; Gruber, 2008; Renkl, 1996) und relativ unabhängig von dem familiären Hintergrund vornehmlich in der Schule erworben werden (Schneider; Küspert & Krajewski, 2013). Sie sind damit ebenfalls recht unabhängig von weiteren Dritteinflüssen und für eine präzise Analyse geeignet. Zum anderen werden Mathematikleistungen gerade aus dem Grund aufgenommen, da bereits zu einem frühen Zeitpunkt vornehmlich die hohe Aufgabenähnlichkeit zwischen kognitiven Testverfahren und mathematischen Leistungen betont wurde (Coley & Lohnes, 1976; Rindermann, 2006, 2008; Ceci, 1991).

Diese hohe Aufgabenähnlichkeit wird bis heute als bedeutsamstes Argument für die Überlappungs- bzw. Identitätsposition angeführt (Rindermann, 2006). Daher soll gerade anhand der vermeintlich ähnlichsten Konstrukte die Differenziertheit geprüft werden.

3.1 Stichprobe

Zur Beantwortung der wissenschaftlich noch ungeklärten Fragestellungen konnte als Datengrundlage des empirischen Teils der vorliegenden Arbeit eine regionale Totalerhebung (Rügen und Rostock) im Rahmen der Mecklenburger Längsschnittstudie der Jahre 2006-2009 herangezogen werden. Die Mecklenburger Längsschnittstudie hat das Ziel, die Merkmale von Kindern mit Entwicklungsrückständen in Leistungs- und Verhaltensbereichen sowie sozialen Aspekten zu ermitteln. Kinder mit potentiell gefährdetem Schulerfolg werden in Mecklenburg teilweise in Diagnoseförderklassen (DFK) oder aber zunehmend inklusiv in regulären Grundschulklassen unterrichtet. Um die Effizienz der Beschulung in Diagnoseförderklassen zu erfassen, wurde im Rahmen der Mecklenburger Längsschnittstudie von Koch, Hartke und Blumenthal (2009) zunächst die Lernausgangslage dieser Kinder festgestellt. Dabei wurden die Lernausgangslagen von Kindern mit vergleichbaren Voraussetzungen, aber unterschiedlicher Beschulungsform erfasst und über die Zeit verfolgt. Um auch die Entwicklung von noch nicht bereits vor der Einschulung diagnostizierten potentiell gefährdeten Kinder beschreiben zu können, wurden nicht nur Kinder, die in Diagnoseförderklassen beschult wurden, untersucht, sondern es wurden die Daten von *allen* Kindern erfasst, die in reguläre Grundschulen eingeschult worden waren. Von den potentiell gefährdeten Kindern wurden zahlreiche weitere Leistungsdaten, motivationale Faktoren, Verhaltensskalen sowie Rechtschreibung und Leseleistungen erhoben. Für weitere Informationen kann auf die Berichte von Koch, Hartke und Blumenthal (2009, 2010) verwiesen werden. Insgesamt wurden die Daten *aller* Kinder aus 37 Schulen und 86 Klassen der Regionen Rügen und Rostock (18 Schulen), die 2006 eingeschult worden sind, in die Gesamtstichprobe der Mecklenburger Längsschnittstudie aufgenommen. Hinsichtlich regionaler Unterschiede konnten bereits im Rahmen des ersten Berichts aus der Mecklenburger Längsschnittstudie keine bedeutenden Unterschiede festgestellt werden (Koch, Hartke & Blumenthal, 2009).

Für die hier relevanten Fragestellungen gingen die Daten aller regulär eingeschulten Kinder der Regionen Rügen und Rostock ein, die sich in den Grundschulen, nicht aber in den Diagnoseförderklassen befanden. Aus der Region Rostock verteilten sich die Kinder auf 18 Schulen mit 45 Grundschulklassen. Aus der Region Rügen waren 17 Schulen mit 27

Grundschulklassen beteiligt. Insgesamt wurden diese Kinder in 35 Schulen und 72 regulären Grundschulklassen beschult. Insofern handelt es sich bei den erhobenen Daten um eine unausgelesene und auch aufgrund des Stichprobenumfangs repräsentative Stichprobe.

Um die im Interesse der längsschnittlichen Zusammenhänge zwischen kognitiven Fähigkeiten und mathematischen Leistungen stehenden Zusammenhänge zu ermitteln, bearbeitete somit eine Stichprobe von $N = 1726$ Kindern an regulären Grundschulen zu vier Messzeitpunkten Aufgaben aus kognitiven Fähigkeitstests für die Klassenstufen 1 bis 4. Als Leistungsmerkmal schulischer Fähigkeiten wurden mathematische Schulleistungen ebenfalls über den Zeitraum von vier Jahren mit vier Messzeitpunkten herangezogen.

Die nachstehenden Tabellen präsentieren die Verteilung der Stichprobe, aufgeschlüsselt nach Region und Geschlecht zu den vier Messzeitpunkten. Die erste Tabelle zeigt die Gesamtübersicht, während die nachfolgende Tabelle die Übersicht der Stichprobe aus Rostocker Grundschulen darstellt, gefolgt von der Übersicht der Stichprobe der Rügener Region. Da weder geschlechtsspezifische Fragestellungen oder regionale Unterschiede im Vordergrund dieser Untersuchung standen (letztere wurden im Rahmen der Mecklenburger Längsschnittstudie von Koch, Hartke und Blumenthal (2009) geprüft und ergaben keine bedeutenden Unterschiede), gingen die Daten von Kindern mit gültigen Eintragungen in den kognitiven Fähigkeitstestverfahren und den Mathematikleistungstests ein, auch wenn diese im Datensatz keinen gültigen Eintrag zu Geschlecht oder regionaler Herkunft hatten. Daher können die folgenden drei Tabellen von den in die Analysen eingegangenen Fallzahlen geringfügig abweichen.

Tabelle 1

Beschreibung der Gesamtstichprobe nach Geschlecht für die Regionen Rostock und Rügen; Umfang der Stichprobe $N = 1487$ (768 Jungen und 719 Mädchen)

	<i>Jungen</i>		<i>Mädchen</i>		<i>Gesamt</i>	
	<i>N</i>	<i>k. A.</i>	<i>N</i>	<i>k. A.</i>	<i>N</i>	<i>k. A.</i>
Messzeitpunkt 1						
KFT 1	740	28	688	31	1428	59
DEMAT 1	617	151	580	139	1197	290
Messzeitpunkt 2						
KFT 2	648	120	575	138	1229	258
DEMAT 2	603	165	528	191	1131	356
Messzeitpunkt 3						
KFT 3	554	214	517	202	1071	416
DEMAT 3	578	190	548	171	1126	361
Messzeitpunkt 4						
KFT 4	534	234	488	231	1022	465
DEMAT 4	489	289	485	234	974	513

N = Anzahl der Kinder mit vollständigen Angaben; $k. A.$ = Anzahl der Kinder mit fehlenden Angaben im Test.

Tabelle 2

Beschreibung der Gesamtstichprobe nach Geschlecht und Region für die Region Rostock; Umfang der Stichprobe $N = 1022$ (524 Jungen und 498 Mädchen)

	<i>Jungen</i>		<i>Mädchen</i>		<i>Gesamt</i>	
	<i>N</i>	<i>k. A.</i>	<i>N</i>	<i>k. A.</i>	<i>N</i>	<i>k. A.</i>
Messzeitpunkt 1						
KFT 1	501	23	473	25	974	48
DEMAT 1	388	136	365	133	753	269
Messzeitpunkt 2						
KFT 2	423	101	382	116	805	217
DEMAT 2	384	140	335	163	719	303
Messzeitpunkt 3						
KFT 3	360	164	347	151	707	315
DEMAT 3	387	137	377	121	764	258
Messzeitpunkt 4						
KFT 4	373	151	349	149	722	300
DEMAT 4	352	172	346	152	698	314

N = Anzahl der Kinder mit vollständigen Angaben; $k. A.$ = Anzahl der Kinder mit fehlenden Angaben im Test.

Tabelle 3

Beschreibung der Gesamtstichprobe nach Geschlecht und Region für die Region Rügen; Umfang der Stichprobe $N = 465$ (244 Jungen und 221 Mädchen)

	<i>Jungen</i>		<i>Mädchen</i>		<i>Gesamt</i>	
	<i>N</i>	<i>k. A.</i>	<i>N</i>	<i>k. A.</i>	<i>N</i>	<i>k. A.</i>
Messzeitpunkt 1						
KFT 1	239	5	215	6	454	11
DEMAT 1	229	15	215	6	444	21
Messzeitpunkt 2						
KFT 2	225	19	199	22	424	41
DEMAT 2	219	25	193	28	412	53
Messzeitpunkt 3						
KFT 3	194	50	170	51	364	101
DEMAT 3	191	53	171	50	362	103
Messzeitpunkt 4						
KFT 4	161	83	139	82	300	165
DEMAT 4	137	117	139	82	266	199

N = Anzahl der Kinder mit vollständigen Angaben; $k. A.$ = Anzahl der Kinder mit fehlenden Angaben im Test.

Der theoriegeleitete Umgang mit fehlenden Werten in einzelnen Messwerten wird darüber hinaus im folgenden Abschnitt berichtet. Von der analysierten Gesamtstichprobe von $N = 1726$ hatten zum ersten Messzeitpunkt 1487 Kinder eine gültige Eintragung zu regionaler Herkunft und Geschlecht. Zum zweiten Messzeitpunkt waren hinsichtlich Geschlecht und Region Eintragungen von 1252 Kindern vorhanden und zum dritten Messzeitpunkt von 1099 Kindern, während diese Informationen zum vierten Messzeitpunkt von 1055 Kindern zur Verfügung standen, anhand derer die Verteilung nach Region und Geschlecht in der obigen Tabelle dargestellt werden kann.

Umgang mit fehlenden Werten

Der angemessene Umgang mit fehlenden Daten erfordert ein theoriegeleitetes Vorgehen. Insbesondere in Längsschnittstudien sind Datenausfälle unvermeidlich. Die Ursache sind Antwortverweigerungen der Testpersonen, unlesbare Antworten oder auch die fehlende Teilnahme an allen Testzeitpunkten. Die resultierenden Probleme sind eine reduzierte Stichprobe und eine geringere Qualität der Schätzwerte (Little & Rubin, 2002, in Lüdtke & Robitzsch, 2010). Zudem erfordert eine Reihe von Analysemethoden vollständige Daten (ebd.). In der Praxis werden daher häufig alle Personen mit fehlenden Werten vollständig ausgeschlossen, oder aber mit dem Mittelwert der Variablen der verbleibenden Personen ersetzt, was allerdings, so Lüdtke und Robitzsch (2010) verzerrte Ergebnisse zur Folge hat. Little und Rubin (1976, in Lüdtke & Robitzsch, 2010) schlugen eine Klassifikation der Entstehung fehlender Werte vor. Demnach werden „vollständig zufällig“ fehlende Werte ohne systematischen Zusammenhang verstanden. Diese haben also keine Verbindung zu anderen Variablen im Datensatz, wie es beispielsweise bei bewusst ausgelassenen Antworten der Fall wäre, wenn eine kritische Variable erhoben werden soll. Weiter gibt es „zufällig“ fehlende Daten, die zwar eine Auswirkung auf andere Variablen besitzen, aber nicht mit der Ausprägung dieser assoziiert sind (Rubin, 1976, in Lüdtke & Robitzsch, 2010). Darüber hinaus gibt es „nicht zufällig“ fehlende Daten, was der Fall ist, wenn Personen bewusst keine Angaben gemacht haben. Die Ursache fehlender Werte lässt sich mit einem Test nach Little (1988, in Lüdtke & Robitzsch, 2010) ermitteln.

Die drei in der Literatur diskutierten Verfahren im Umgang mit fehlenden Daten sind klassische Verfahren, die einen fallweisen Ausschluss der Daten aus der Analyse zur Folge haben. Weiter gibt es die Möglichkeit, Daten zu ersetzen, sogenannte imputationsbasierte und modellbasierte Verfahren, die gleichzeitig die Schätzung des Modells und fehlende Daten berücksichtigen (Lüdtke & Robitzsch, 2010).

In der vorliegenden Arbeit wurde die Ursache fehlender Werte überprüft und nur einzelne fehlende Daten innerhalb der Subskalen der relevanten Testverfahren zur Erfassung der kognitiven Fähigkeiten und mathematischer Schulleistungen, aus der Analyse ausgeschlossen, da die Personen zumeist nur einzelne Items ausgelassen haben, die keinen systematischen Zusammenhang zu anderen Variablen aufwiesen. Personen, die nur an einem oder zwei Messzeitpunkten gefehlt haben, wurden in der Analyse aufrechterhalten und die Angaben zu den Messzeitpunkten, zu denen Daten vorhanden waren, erfasst und in die Ana-

lysen einbezogen. So erklären sich die geringen Abweichungen der für die Analysen einbezogenen Fallzahlen. Damit wurde die Stichprobe so groß wie möglich gehalten, die Anzahl der verbleibenden Items einer Subskala ebenso, ohne aber unnötig auf verwertbare Daten zu verzichten oder durch Imputation eine verzerrte Parameterschätzung zu riskieren.

In der vorliegenden Untersuchung wurden nur vollständig fehlende Angaben der relevanten Testergebnisse der kognitiven Fähigkeiten und mathematischen Schulleistungen aus der Analyse ausgeschlossen, wie in der nachfolgenden Tabelle dargestellt.

Tabelle 4

Übersicht über die den einzelnen Fragestellungen zugrundeliegenden Fallzahlen; Umfang der zugrundeliegenden Fälle ($N = 1726$)

Fragestellung	MZP	analysierte Fälle	ausgeschlossene Fälle	Output auf beiliegender CD
1		1576	150	Messinvarianz
2		1700	26	Autoregressive Effekte
3		1700	26	Kreuzverzögerte Effekte
4	1	1499	227	Faktorielle Differenziertheit
	2	1396	330	
	3	1402	324	
	4	1336	390	

MZP = Messzeitpunkt.

3.2 Studiendesign

Zur Beantwortung der wissenschaftlich noch ungeklärten Fragestellungen wurde als Datengrundlage des empirischen Teils die oben beschriebene, repräsentative, unausgelesene Stichprobe herangezogen, die sich in der ersten bis vierten Klassenstufe regulärer Grundschulen befanden. Die folgende Abbildung zeigt das Studiendesign in der Übersicht.

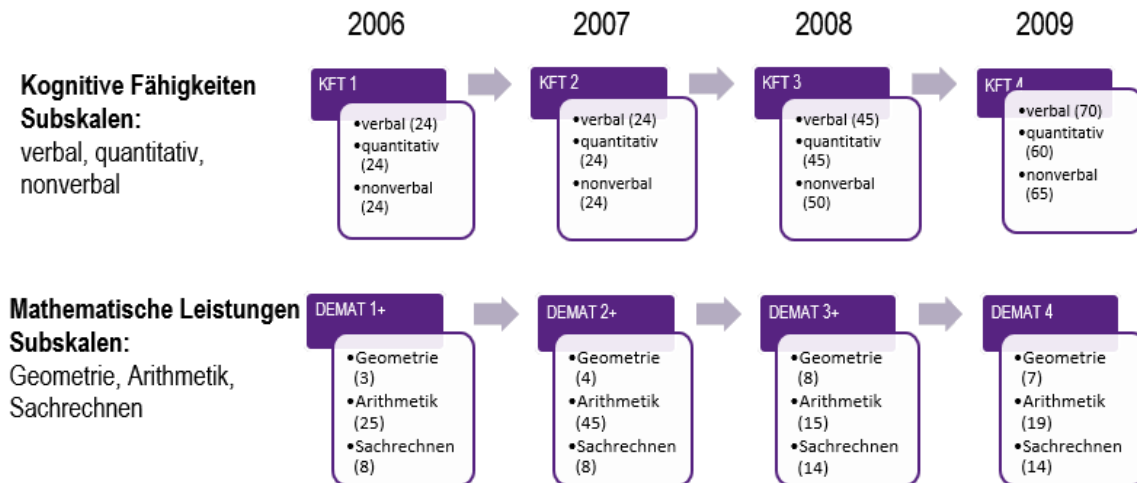


Abbildung 10 Übersicht über die durchgeführten Datenerhebungen von 2006 bis 2009.

In der oberen Zeile sind die eingesetzten Messinstrumente zur Erfassung der kognitiven Fähigkeiten der KFT-Reihe für die erste, zweite, dritte Klassenstufe, sowie für die vierte Klassenstufe von Abou-Koura und Perleth (2005), Perleth und Heller (2008) und Heller und Perleth (2000) dargestellt. Zur Erfassung der mathematischen Schulleistungen wurden die Testverfahren der DEMAT-Reihe (Deutscher Mathematiktest für erste, zweite, dritte und vierte Klassenstufen) von Krajewski, Küspert und Schneider (2002), Krajewski, Liehm und Schneider (2004) sowie Roick, Gölitze und Hasselhorn (2004, 2006) eingesetzt. Die Testverfahren für die kognitiven Fähigkeiten sind in die jeweiligen Subtests: verbale, quantitative und nonverbale Fähigkeiten unterteilt. In Klammern ist die Anzahl der Items der Subtests angegeben.

Die untere Zeile zeigt die Testinstrumente der Mathematikleistungen zu den vier Messzeitpunkten im jährlichen Abstand. Die Testverfahren Deutsche Mathematiktests der DEMAT-Reihe sind in die Subtests Geometrie, Arithmetik und Sachrechnen nach Stoffgebieten klassifiziert. Die Anzahl der Items ist ebenso in Klammern angegeben. Die eingesetzten Messinstrumente werden im folgenden Kapitel näher erläutert.

3.3 Messinstrumente

3.3.1 Kognitive Fähigkeitstestverfahren der KFT – Reihe

Im Laufe der Forschungsentwicklung setzten sich im Einklang mit der Entwicklung der relevanten Theorien Tests durch, die die bereichsspezifische und hierarchische Struktur der Intelligenz voraussetzen. Mit dem Einsatz der Testverfahren der kognitiven Fähigkeitstests

der KFT-Reihe von Heller und Perleth (2000), Perleth und Heller (2008) und Abou-Koura und Perleth (2005) wird dieser Bereichsspezifität in der vorliegenden Arbeit Rechnung getragen. Es werden drei Dimensionen kognitiver Fähigkeiten unterschieden: verbale, quantitative und nonverbale. Eine allgemeine kognitive Komponente, die stabile allgemeine Intelligenz, beeinflusst alle spezifischen Dimensionen in gleicher Weise. Es werden im KFT-Modell sowohl allgemeine als auch spezifische Anteile berücksichtigt. Die Entwicklung und die Konzeption des Testverfahrens basieren auf dem Cognitive-Ability-Test von Thorndike und Hagen (1978) und ermöglichen eine differenzierte Erfassung von Fähigkeitsdimensionen. Es geht zurück auf das Primärfaktorenmodell von Thurstone (1938).

Mittels der quantitativen Dimension werden Textrechenaufgaben, arithmetisches Denken und Mengenvergleiche als Subskalen erfasst. Weiter sind die Bildung von Zahlenreihen, Rechenfertigkeiten und das Bilden von Gleichungen erforderlich. Innerhalb der nonverbalen Kategorie sind Fähigkeiten zur Klassifikationen von Figuren, anschauungsgebundenes Denken und die Fähigkeit zur Bildung von Figurenanalogien erforderlich und auch konstruktive Fähigkeiten (Figurensynthese) im nonverbalen Teil gefragt. Die KFT-Reihe ist als Paper-Pencil-Version als Gruppentest- und als Einzeltest einsetzbar. Die Erfassungsdimensionen des KFT orientieren sich an verschiedenen Intelligenzmodellen, wie dem Primärfaktorenmodell nach Thurstone (1938) und dem Berliner Intelligenzstrukturmodell von Jäger et al. (1997).

Für die Erfassung der kognitiven Fähigkeiten in der ersten und zweiten Klassenstufe wurden jeweils die Revisionsversionen des KFT 1-2 R (Abou-Koura & Perleth, 2005) eingesetzt. Das Verfahren des KFT 3R von Perleth und Heller aus dem Jahr 2008 diene der Messung kognitiver Fähigkeiten, die in der dritten Klassenstufe erbracht wurden, während das Testverfahren KFT 4-12 R von Heller und Perleth aus dem Jahr 2000 zur Erfassung der Leistungen in der vierten Klassenstufe eingesetzt wurde.

3.3.1.1 KFT 1-2 R

Der kognitive Fähigkeitstest für die ersten und zweiten Klassenstufen (KFT 1-2 R, Abou-Koura & Perleth, 2005) ist ein Testverfahren zur Messung kognitiver differenzierter Fähigkeitsdimensionen von Schülerinnen und Schülern für die ersten und zweiten Klassenstufen.

Der KFT 1-2 R ist ein reiner Power-Niveau-Test ohne Zeitbeschränkungen. Als Zeitorientierung werden von den Autoren jedoch für die Bearbeitung des verbalen Teils 18 (Untertest 1) und 15 Minuten (Untertest 2) angegeben. Für den quantitativen Teil sind 18 Minuten für den Untertest *Beziehungen erkennen* und 17 Minuten für das *Rechnerische*

Denken vorgesehen. Für die Bearbeitung des nonverbalen Teils stehen 14 Minuten für das *Schlussfolgernde Denken* und 16 Minuten für die Bearbeitung des Untertests *Matrizen* zur Verfügung.

Die Testaufgaben beruhen auf bildlichen Vorlagen, zu denen die Instruktionen mündlich gegeben werden. Die Schüler markieren die richtigen Lösungen in ihrem Testheft.

Die Dimensionen *verbale Fähigkeiten* sind in 2 Subtests, *Sprachverständnis* (V1) und *Wortschatz* (V2) unterschieden, die die entsprechenden Fähigkeiten jeweils mittels 24 Items erfassen. Der Test misst die Fähigkeit zum induktiven Reasoning, die Problemlösefähigkeit und das Sprachverständnis. Es sollen anhand der bildlichen Vorlagen Gegenstände oder Handlungen benannt bzw. die Handlung bei Vorgabe ihrer Verwendung erkannt werden. Es soll also geschlussfolgert oder auch Synonyme für vorgegebene Wörter gefunden werden. Im Subtest V2 sollen zudem Satzergänzungen vorgenommen werden.

Die *quantitative Fähigkeitsdimension* besteht aus den Subtests *Beziehungen erkennen* (Q1) und *Rechnerisches Denken* (Q2). Es werden allgemeine abstrakte Reasoning-Fähigkeiten, vor allem induktives Reasoning, gefordert. An dieser Stelle wird der Umgang mit Zahlbegriffen und mengenmäßigen Vorstellungen erfasst. Der Aufgabentyp von Test Q2 sei an folgendem Beispiel erläutert:

„Schaut auf das erste Bild in Reihe 1. Zählt die Vögel, die auf dem Zaun sitzen. Vorher saßen noch zwei weitere Vögel auf dem Zaun, dann flogen zwei Vögel weg. Füllt den Kreis unter dem Bild aus, das zeigt, wie viele Vögel auf dem Zaun zurückblieben. Denkt daran, zwei weitere Vögel saßen vorher auf dem Zaun und dann flogen zwei weg.“

(Abou-Koura & Perleth, 2005)

In dem Subtest *Beziehungen erkennen* sollen Beziehungskonzepte erkannt oder logische Folgen identifiziert werden. Die Lösung der Aufgaben erfordert die Identifikation von Größen, Standorten oder Mengen. Es müssen zeitliche, räumliche und größenmäßige Beziehungen aus den bildlichen Vorlagen erkannt werden. Der Aufgabentyp *Beziehungen erkennen* von Test Q1 wird an folgendem Beispiel erläutert:

„Schaut Euch die Kinder in Reihe 1 an. Findet das Kind, das sowohl groß als auch dünn ist. Füllt den Kreis unter diesem Kind aus.“ (Abou-Koura & Perleth, 2005)

In der Subskala *Rechnerisches Denken* werden einfache Reihenprobleme erfragt, die Kinder mit grundlegenden mathematischen Regeln lösen sollen.

In der *nonverbalen Fähigkeitsdimension* werden *Schlussfolgerndes Denken* (N1) und *Matrizen* (N2) unterschieden. Die Skala *Schlussfolgerndes Denken* erfasst vornehmlich induktive Reasoning-Fähigkeiten. Dabei sollen Grundregeln generalisiert und transferiert werden. In der Skala *Schlussfolgerndes Denken* sollen Ähnlichkeiten in den Abbildungen erkannt und aus vier Antwortabbildungen die logische Ergänzung ausgewählt werden, wie die folgende Abbildung auszugsweise veranschaulicht. Die Schülerinnen und Schüler müssen erkennen, unter welchem Aspekt sich jeweils vier von fünf vorgegebenen Bildervorlagen gleichen. Anschließend ist die jeweils nicht dazu passende Abbildung anzukreuzen.

In der Matrizenkala ist zu jedem Item ein Quadrat mit vier Kategorien abgebildet. Drei Kästchen enthalten Abbildungen, das vierte ist frei und in dieses soll aus den separat angegebenen Antwortmöglichkeiten die richtige Abbildung gewählt werden.

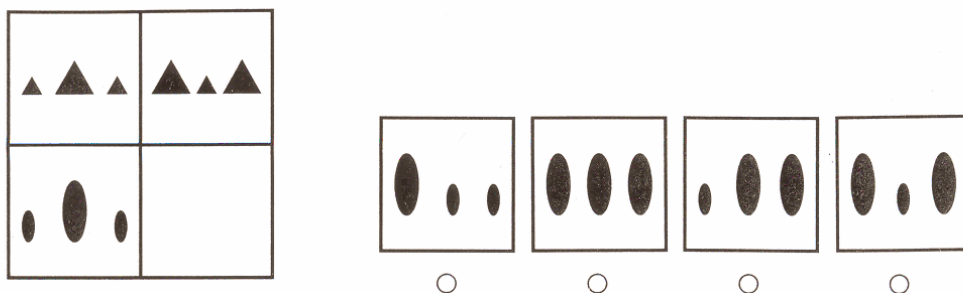


Abbildung 11 Beispiel Matrizen N2.

Den Kindern wird ein großes Quadrat vorgegeben, welches in vier kleinere Kästchen unterteilt ist. Mit einem der Teile sollen sie das leere Feld vervollständigen. Die Kinder sollen die Beziehung zwischen den drei gegebenen Kästchen herausfinden. Im folgenden Kapitel wird das Testverfahren zur Erfassung kognitiver Fähigkeiten für die dritte Klassenstufe vorgestellt.

3.3.1.2 KFT 3 R

Der kognitive Fähigkeitstest für die dritte Klassenstufe von Kawthar Abou-Koura und Christoph Perleth wurde 2005 auf der Basis des kognitiven Fähigkeitstests für das Kindergartenalter (KFT-K, 1983) und des kognitiven Fähigkeitstests für die Grundschule zur Ver-

wendung in den ersten bis dritten Klassenstufen (KFT 1-3) von Heller und Geisler konzipiert. Beide zugrundeliegenden Testversionen finden ihren Ursprung in den Parallelformen des Primary-Tests I und II sowie der Cognitive Ability Tests (Thorndike & Hagen, 1978).

Der Testaufbau unterscheidet, wie auch das zuvor beschriebene Testverfahren des KFT 1-2 R, drei Dimensionen. Diese sind in der Version KFT 3 R in je zwei Subtests gegliedert: Im Bereich Sprachverständnis (siehe Beispiel in der folgenden Abbildung) erfolgt die Unterteilung in den verbalen Untertest (V1) und den verbalen Untertest (V2). Im V1-Teil wird der Wortschatz erfasst. Die Kinder sollen zu vorgegebenen Wörtern Synonyme oder Oberbegriffe aus jeweils fünf Auswahlmöglichkeiten markieren.

Beispiel zu V1:

Rose *A: Musik* *B: Blume* *C: Speise* *D: Aussicht* *E: Lasso*

Beispiel zu V2:

Das Feuer ist _____

A: feucht *B: grün* *C: heiß* *D: flüssig* *E: kräftig*

In der quantitativen Dimension werden im Q1-Subtest Mengenvergleichsfähigkeiten erfasst, während in Q2 Zahlenreihen logisch vervollständigt werden sollen, wie folgendes Beispiel zeigt.

Beispiel zu Q2:

1 2 3 + + → A: 3 B: 4 C: 5 D: 6 E: 8

Bei dieser Aufgabe sollen Ziffern und Operationszeichen so zusammengestellt werden, dass jeweils eine richtige Rechnung (Gleichung) entsteht. Auf dem Auswahlantwortbogen ist dann die korrekte Lösung anzukreuzen (hier Lösung D). Die nonverbale Dimension ist ebenfalls in einen N1- und einen N2-Teil unterteilt. Im ersten Untertest wird die Fähigkeit zur Figurenklassifikation ermittelt, indem aus fünf Auswahlmöglichkeiten diejenige markiert werden soll, die zu den drei vorgegebenen Figuren passt. Im N2-Untertest sollen dann Figurenpaare vervollständigt werden. Während die Subtests *Sprachverständnis* und *Quan-*

titatives Denken die schulische Lernfähigkeit erfassen, wird in der Kategorie *Schlussfolgerndes Denken* die allgemeine Denkfähigkeit erfasst, die als Indikator für das allgemeine kognitive Niveau gilt.

Die Durchführung kann ebenso wie das zuerst beschriebene Verfahren als Einzel- oder als Gruppentest erfolgen. Es handelt sich um einen kombinierten Power-Speed-Test in einer Paper-Pencil-Version. Als Zeitorientierungsrahmen sind für V1 und V2 je 7 Minuten, für Q1 10 Minuten, Q2 und N1 9 Minuten und für N2 8 Minuten vorgesehen. Nachfolgend wird das Testverfahren für die vierte Klassenstufe erläutert.

3.3.1.3 KFT 4-12 R

Der kognitive Fähigkeitstest für die vierte bis zwölfte Klassenstufe (KFT 4-12 R) von Kurt Heller und Christoph Perleth (Heller & Perleth, 2000) erschien im Jahr 1999 als Revisionsform auf die Testversion von Heller, Gaedicke und Weinländer aus dem Jahr 1976. Im Vergleich zur Vorversion bietet die Revisionsform erweiterte Anwendungsmöglichkeiten und Detailverbesserungen hinsichtlich der Validität (Heller & Perleth, 2000). Der Testaufbau entspricht den zuvor beschriebenen Testverfahren mit einer Unterscheidung in drei Testteildimensionen (*verbal, quantitativ und nonverbal*). Die Differenzierung der *verbalen Dimension* erfolgt in drei Kategorien V1, V2 und V3. Im Verbalteil 1 (V1) wird Wortschatzwissen erfragt, indem zu vorgegebenen Wörtern Synonyme bzw. Oberbegriffe gefunden und auf dem Antwortblatt markiert werden müssen. Im Verbaltest V2 dagegen wird die Fähigkeit zur Wortklassifikation gefragt, indem Wörter der gleichen Wortklasse bzw. des gleichen Oberbegriffs zu drei vorgegebenen Worten gefunden werden sollen. Im V3-Untertest werden Fähigkeiten zur Bildung von Wortanalogien gemessen, indem jeweils ein Wortpaar, das in einem Zusammenhang steht, zu einem dritten Wort genau das Lösungswort aus fünf Antwortmöglichkeiten ergänzt werden sollen, das einen äquivalenten Zusammenhang zu dem dritten vorgegebenen Wort darstellt, sodass ein neues Wortpaar gebildet wird.

Die *quantitative Dimension* besteht ebenfalls aus drei Untertests (Q1, Q2 und Q3). Im ersten Untertest werden Mengenvergleichsfähigkeiten erfragt, indem unter drei Optionen die Entscheidung verlangt wird, ob die beiden vorgegebenen Mengen gleich groß oder eine Menge kleiner bzw. größer ist. Im zweiten Untertest (Q2) sollen Zahlenreihen gelöst werden, indem unter den Antwortoptionen die sinnvolle Ergänzung markiert werden soll. Im dritten Untertest sollen zu vorgegebenen Zahlen angemessene Operationszeichen so ausgewählt und eingesetzt werden, dass als Resultat eines der vorgegebenen Auswahlresultate korrekt ist.

Die *nonverbale Dimension* (N-Teil) des KFT 4-12 R ist ebenfalls in drei Untertests (N1, N2 und N3) unterteilt. Im N1-Untertest werden Fähigkeiten zur Figurenklassifikation erfasst, indem zu drei vorgegebenen Figuren, die ein Merkmal gemeinsam haben und damit einer Klasse angehören, eine weitere unter fünf Auswahlmöglichkeiten an Figuren ergänzt und markiert werden muss. Im N2-Untertest hingegen wird die Fähigkeit zur Bildung von Figurenanalogien ermittelt, indem Figurenpaare vervollständigt werden sollen. Der N3-Untertest imitiert Faltaufgaben. Er verlangt von den Kindern die mentale Vorstellung, ein Blatt Papier gedanklich zu falten, zu lochen, um dann eine Auswahlalternative zu markieren, die sich nach dem Auseinanderfalten ergeben würde.

Der Test wird als kombinierter Power-Speed-Test mit ansteigender Schwierigkeit und einer reinen Bearbeitungszeit von 80 Minuten durchgeführt. Er sieht jedoch nach der Bearbeitung des jeweiligen verbalen, quantitativen und nonverbalen Teils Pausen vor und liegt in den Parallellformen A und B vor. Die Auswertung kann mittels Lösungsschablonen oder computergestützt erfolgen. Sie erfolgt entsprechend der Untertests und ergibt ein Profil der kognitiven Fähigkeiten, wobei zudem die Ermittlung eines Gesamtleistungswertes möglich ist.

Nähere Informationen zu den Items im Einzelnen, sowie zu den Gütekriterien des KFT 1, 2 und 3 können mit dem Hinweis auf die Kontaktierung der Testautoren (Perleth & Heller) erfragt werden. Die Detailinformationen zum KFT 4-12 sind unter Hinweis auf Urheberrechte beim Hogrefe Verlag GmbH & Co. KG, Göttingen zu erhalten. Nähere Informationen zu den Items im Einzelnen, sowie zu den Gütekriterien des KFT 1, 2 und 3 können mit dem Hinweis auf die Kontaktierung der Testautoren (Perleth und Heller) erfragt werden. Die Detailinformationen zum KFT 4-12 sind unter Hinweis auf Urheberrechte beim Hogrefe Verlag GmbH & Co. KG, Göttingen zu erhalten. Im folgenden Kapitel werden die Testverfahren zur Erfassung mathematischer Schulleistungen vorgestellt.

3.3.2 Mathematische Schulleistungstests der DEMAT-Reihe

Für die Leistungserfassung der Mathematikleistungsbereiche wurden die jeweiligen Schulleistungstests in Mathematik der Reihe Deutscher Mathematikleistungstest (DEMAT-Versionen) für die ersten, zweiten, dritten sowie in der vierten Klassenstufe von Krajewski, Küspert und Schneider aus dem Jahr 2002, Krajewski, Liehm und Schneider aus dem Jahr 2004 und von Roick, Gölitze und Hasselhorn aus den Jahren 2004 und 2006 eingesetzt.

Diesem Testverfahren liegt das Stufenmodell des Erwerbs arithmetischer Operationen nach Aebli (1976) zugrunde. Es basiert auf der Annahme, dass jede arithmetische Operation eine abstrakte Handlung ist (Krajewski, Küspert & Schneider, 2002), die mentale Ansatzbildung einer mathematischen Operation beschreibt, die zu einem Verständnis der abstrahierten Rechenoperation führt. Es erfordert Vorstellungsfähigkeit von Handlungen und mentalen Rückgriff auf Ziffern, deren Darstellung oder bildlichen Entsprechung, welches schließlich dazu führt, deklaratives sowie prozedurales Wissen aufzubauen (Krajewski, Küspert & Schneider, 2002).

Die Testverfahren der DEMAT-Reihe folgen der Unterscheidung nach Stoffgebieten, wie sie beispielsweise von Brunner (2006) begründet wurde. Der Mathematikleistungstest der DEMAT-Reihe unterscheidet mathematische Teilbereiche auch mit der Begründung, dass pädagogische Interventionen auf die entsprechenden Problembereiche der einzelnen Schüler ausgerichtet werden können (Krajewski, Küspert & Schneider, 2002).

In der unten abgebildeten Tabelle wird eine Übersicht über die Items der vier Messverfahren für die erste bis vierte Klassenstufe der DEMAT-Testverfahrenreihe von Krajewski, Küspert und Schneider (2002), Krajewski, Liehm und Schneider (2004) sowie Roick, Gölitze und Hasselhorn (2004, 2006) dargestellt.

In dem Schulleistungstest für die erste Klassenstufe werden die Kategorien *Mengen-Zahlen* (3 Items), *Zahlenraum* (5 Items), *Addition* (4 Items), *Subtraktion* (4 Items), *Zahlenzerlegung-Zahlenergänzung* (4 Items), *Teil-Ganzes-Aufgaben* (4 Items), *Kettenaufgaben* (4 Items) und *Ungleichungen* (4 Items) sowie *Sachrechenaufgaben* (4 Items) differenziert erhoben.

In dem Testverfahren der DEMAT-Reihe für die zweite Klassenstufe werden die Aufgabenbereiche in *Zahleneigenschaften* (2 Items), *Längenvergleich* (4 Items), *Addition* (4 Items), *Subtraktion* (4 Items), *Verdoppeln* (3 Items), *Division* (4 Items), *Halbieren* (3 Items), *Rechnen mit Geld* (4 Items) und *Sachrechenaufgaben* (4 Items) sowie *Geometrie* (4 Items) unterschieden.

Die Unterscheidung in die Stoffgebiete *Geometrie*, *Sachrechnen* und *Arithmetik* erfolgt entsprechend den drei in der Grundschulmathematik bedeutsamen Kompetenzbereichen (Krajewski, Küspert & Schneider, 2002). Die untenstehende Tabelle zeigt eine Übersicht über die Skalen und zugrundeliegende Itemanzahl für die Verfahren je Klassenstufe.

Tabelle 5

Übersicht über Skalen und Items zur Messung mathematischer Schulleistungen aus der Reihe Deutscher Mathematiktests (DEMAT) 1. bis 4. Klassenstufe

DEMAT 1		DEMAT 2		DEMAT 3		DEMAT 4	
<i>Skala</i>	<i>N</i>	<i>Skala</i>	<i>N</i>	<i>Skala</i>	<i>N</i>	<i>Skala</i>	<i>N</i>
<i>MZ</i>	3	<i>ZE</i>	2	<i>ZS</i>	3	<i>ZS</i>	3
<i>ZR</i>	5	<i>LÄ</i>	4	<i>AD</i>	4	<i>AD</i>	4
<i>AD</i>	4	<i>AD</i>	4	<i>SU</i>	4	<i>SU</i>	4
<i>SU</i>	4	<i>SU</i>	4	<i>MU</i>	4	<i>MU</i>	4
<i>ZZ</i>	4	<i>DO</i>	3	<i>SR</i>	4	<i>DI</i>	4
<i>TG</i>	4	<i>DI</i>	4	<i>SZ</i>	3	<i>GV</i>	6
<i>KA</i>	4	<i>HA</i>	3	<i>FL</i>	3	<i>SR</i>	8
<i>UG</i>	4	<i>GEL</i>	4	<i>LS</i>	2	<i>LB</i>	4
<i>SA</i>	4	<i>SA</i>	4	<i>LU</i>	4	<i>SZ</i>	3
		<i>GEO</i>	4				

N = Anzahl der Items; *MZ* = Mengen-Zahlen; *ZR* = Zahlenraum; *AD* = Addition; *SU* = Subtraktion; *ZZ* = Zahlenzerlegung-Zahlenergänzung; *TG* = Teil-Ganzes; *KA* = Kettenaufgaben; *UG* = Ungleichungen; *SA* = Sachaufgaben; *ZE* = Zahleneigenschaften; *LÄ* = Längenvergleich; *DO* = Verdoppeln; *DI* = Division; *HA* = Halbieren; *GEL* = Rechnen mit Geld; *GEO* = Geometrie; *ZS* = Zahlenstrahl; *MU* = Multiplikation; *SR* = Sachrechnen; *SZ* = Spiegelzeichnungen; *FL* = Formen legen; *LS* = Längen schätzen; *LU* = Längen umrechnen; *DI* = Division; *GV* = Größenvergleiche; *LB* = Lagebeziehungen; *SZ* = Spiegelzeichnungen.

In dem Testverfahren für die dritte Klassenstufe erfolgt die Unterscheidung der drei übergeordneten Bereiche in *Arithmetik mit Zahlenstrahlen* (3 Items), *Addition* (4 Items), *Subtraktion* (4 Items), *Multiplikation* (4 Items), dann *Sachrechnen* aus *Sachrechnungen* (4 Items) und *Längen umrechnen* (4 Items), sowie *Geometrie* mit *Spiegelzeichnungen* (3 Items), *Formen legen* (3 Items) und *Längen schätzen* (2 Items) als Teilaufgaben.

Die *Arithmetikaufgaben* der vierten Klassenstufe beinhalten *Zahlenstrahlaufgaben* (3 Items), *Additions-* (4 Items) und *Subtraktionsaufgaben* (4 Items), *Multiplikationsaufgaben* (4 Items) und *Division* (4 Items). Aus dem Bereich *Sachrechnen* werden *Größenvergleiche* (6 Items) und *Sachrechnungen* (8 Items) als Subtestaufgaben erfasst und aus dem

Bereich *Geometrie* die *Lagebeziehungen* (4 Items) und *Spiegelzeichnungen* (3 Items) als Subtestaufgaben.

Die Tests der DEMAT-Reihe sind sowohl als Gruppentest als auch als Einzeltest als Paper-Pencil-Test in einer begrenzten Zeit durchzuführen. Damit ist er für Aussagen zur sozialen sowie kriterialen Bezugsnorm geeignet. Nähere Informationen zu den Items im Einzelnen, können mit dem Hinweis auf Urheberrechte beim Hogrefe Verlag GmbH & Co. KG, Göttingen erhalten werden. Im Folgenden wird auf die Analysemethoden zur Beantwortung der einzelnen Fragestellungen eingegangen.

3.4 Analysemethoden

Zur Analyse der längsschnittlichen Daten der vorliegenden Studie werden Strukturgleichungsmodelle eingesetzt. Sie bieten den Vorteil, dass sich mit ihnen komplexe Zusammenhänge modellieren lassen. Dabei ist es über die Spezifikation von Latent-State-Modellen möglich, Messfehlereinflüsse zu berücksichtigen. Die Parameterschätzung erlaubt, dass die Diskrepanz der modellimplizierten Zusammenhänge (Kovarianz) und den beobachteten Zusammenhängen minimal ist (Wagner, 2014). In der Modelltestung wird geprüft, inwieweit die erwarteten Varianzen und Kovarianzen mit den ermittelbaren Varianzen und Kovarianzen übereinstimmen. Die Latent-State-Trait-Theorie basiert auf der Überlegung von Yousfi und Steyer (2006), dass Testergebnisse nie unabhängig von systematischen Situationseinflüssen entstehen. Diese werden explizit in einem Latent-State-Modell berücksichtigt. So kann die Zusammenhangsstruktur zwischen den latenten Variablen korrelativ beschrieben werden.

Es wird theoriegeleitet entschieden, inwieweit Variablen durch andere Variablen Zusammenhänge aufzeigen könnten (Perels, Otto & Schmitz, 2008), welche Variablen als abhängige betrachtet werden und welche als unabhängige angesehen werden. Die Zusammenhänge werden getestet, um vor allem den Grad der Beeinflussung abzuschätzen. Um jedoch in komplexen Strukturgleichungsmodellen situationsspezifische, also messungsbedingte Schwankungen von tatsächlichen Veränderungen trennen zu können, werden Latent-State-Modelle als geeignete Analysemodelle genutzt. Mit ihnen kann berücksichtigt werden, inwieweit nicht direkt beobachtete Einflüsse latenter Variablen Einfluss auf gemessene Indikatoren nehmen. Latente Variablen können untereinander korreliert sein. Eine hohe Korrelation der Latent-State-Variablen spricht für eine hohe Stabilität interindividueller Unterschiede (Geiser, 2011). Die Reliabilität als Zuverlässigkeitsmaß der Messpräzision gibt

Auskunft über das Varianzverhältnis zwischen beobachteter Testleistung und latenter Leistungsfähigkeit. Im Sinne der klassischen Testtheorie kann ein Vergleich der Reliabilitätskoeffizienten sowohl Informationen zur Stabilität des gemessenen Merkmals liefern, als auch zur Äquivalenz verschiedener Testverfahren sowie zur Invarianz der Testaufgaben (Eid, Gollwitzer & Schmidt, 2013).

Zur Beantwortung der Fragestellung 1 werden die jeweiligen Messmodelle eines Zeitpunktes zu einem Strukturmodell verbunden, sodass der Frage der partiellen Unabhängigkeit der Konstrukte (separat für jeden Messzeitpunkt) nachgegangen werden kann.

Es muss sichergestellt werden, dass in den Verfahren vergleichbare Konstrukte repräsentiert werden und diese somit eine vergleichbare Messstruktur aufweisen (Geiser, 2011). Die zentrale Frage nach der Vergleichbarkeit kann mit der Testung der Messinvarianz beantwortet werden. Mit der Testung der Messinvarianz werden folglich die Indikatoren (beobachtete Variablen) mit den latenten Variablen (nicht beobachteten Variablen) verknüpft und auf ihre zeitliche Konstanz der Parameter hin getestet. Diese Parameter sind Faktorladungen (Regressionen), Intercepts (Mittelwertstrukturen) und Messfehlervarianzen. Wenn die faktoriellen Invarianzbedingungen zufriedenstellend sind, beruhen Werte latenter Variablen zu jedem Zeitpunkt der Messung auf der gleichen Metrik und spätere Schlussfolgerungen sind gerechtfertigt (Widaman, Ferrer und Conger, 2010). Voraussetzung für die Verknüpfung der Messmodelle mit einem Strukturgleichungsmodell ist das Vorhandensein mehrerer Indikatoren für jeden Messzeitpunkt (Geiser, 2011). Erst dann lassen sich Korrelationen der State-Faktoren als Stabilitätsindikatoren hinsichtlich interindividueller Unterschiede eines Konstrukts interpretieren (Geiser, 2011). In der vorliegenden Untersuchung kann aufgrund des Vorhandenseins mehrerer latenter Variablen ein Strukturgleichungsmodell entworfen werden, in dem zusätzlich die Invarianzen der Pfade, der Korrelationen zwischen den latenten Variablen, der Varianzen, der Mittelwerte und der Residuen ermittelbar sind, wodurch eine Prüfung der strukturellen Invarianz möglich wird. Im Messmodell wird spezifiziert, inwieweit latente Variablen (Faktoren) durch beobachtete (manifeste Indikatoren) gemessen werden. Im Anschluss wird das auf die Datenstruktur am besten passende Messmodell (MM) mit dem Strukturmodell (LS) zu einem Strukturgleichungsmodell (ARM) verknüpft (Eid, Gollwitzer & Schmitt, 2013).

Für die Beantwortung der Frage 2 wird je ein Messmodell für die gemessenen Konstrukte kognitive Fähigkeiten (mittels KFT) und mathematische Leistungen (mittels DEMAT) spezifiziert. Auf latenter Ebene werden die Parameter berücksichtigt und in einem zweiten Schritt werden die Parameter aller vier Messzeitpunkte zu einem Strukturmodell

verbunden. Dabei werden die Faktorladungen, Residualkovarianzen und Residualkorrelationen eines Messzeitpunktes ermittelt. Die Spezifikation für das Konstrukt mathematische Schulleistungen erfolgt analog. Die Parameterschätzung erlaubt, dass die Diskrepanz der modellimplizierten Zusammenhänge (Kovarianz) und den beobachteten Zusammenhängen minimal ist (Wagner, 2014). In der Modelltestung wird geprüft, inwieweit die erwarteten Varianzen und Kovarianzen mit den ermittelbaren Varianzen und Kovarianzen übereinstimmen.

Für die Erfassung der Veränderungen im zeitlichen Längsschnitt zur Beantwortung der dritten Fragestellung werden im Folgeschritt über eine, die latenten Einflüsse berücksichtigende Pfadanalyse, zunächst autoregressive Strukturgleichungsmodelle spezifiziert (ARM). Dies geschieht zunächst getrennt für die Konstrukte kognitive Fähigkeiten und mathematische Schulleistungen, indem die Regressionen von einem zum nachfolgenden Zeitpunkt ermittelt werden.

Tabelle 6

Übersicht über die Teilstichproben zur Beantwortung der Fragestellungen; Umfang der zugrundeliegenden Fälle $N = 1726$

Fragestellung	Datengrundlage	analysierte Fälle	fehlende Fälle	Output auf beiliegender CD
1	KFT 1 & DEMAT 1	1499	227	Faktorielle Differenziertheit
	KFT 2 & DEMAT 2	1396	330	
	KFT 3 & DEMAT 3	1402	324	
	KFT 4 & DEMAT 4	1336	390	
2	KFT 1 bis 4 & DEMAT 1 bis 4	1576	150	Messinvarianz
3		1700	26	Autoregressive Effekte
4		1700	26	Kreuzverzögerte Effekte

Für die Erfassung im Längsschnitt werden dann die Regressionen auf zweiter Ordnungsebene des zugrundeliegenden Zeitpunktes das Konstrukt auf den übernächsten Messzeitpunkt und dann auf dritter Ordnungsebene über zwei Messzeitpunkte hinweg ermittelt. Im

Anschluss wird dieses Modell dann ein Modell mit kreuzverzögerten Effekten überführt und miteinander verbunden, um den Grad der gegenseitigen Beeinflussung (Fragestellung 4) abzuschätzen. Die Tabelle 6 zeigt die Übersicht über die zur Beantwortung der jeweiligen Fragestellungen eingehenden Fälle. Die erhobenen Daten werden mit den statistischen Softwarepaketen SPSS 20 und Mplus 7.1 ausgewertet (Muthén & Muthén, 2010). Die methodische Beschreibung der Analysestrategien erfolgt in den folgenden Abschnitten und gliedert sich nach den entsprechenden Fragestellungen.

3.4.1 Analysestrategie 1 – Faktorielle Differenziertheit

Für die Unabhängigkeitsprüfung der Messverfahren der Konstrukte kognitive Fähigkeiten und mathematische Schulleistungen für die jeweiligen Messzeitpunkte bzw. hier Klassenstufen, werden die Korrelationen zunächst auf der manifesten Ebene, dann auf latenter Ebene geprüft. Dafür werden spezifizierte Latent-State-Messmodelle der kognitiven Testverfahren und der Testverfahren zur Erfassung mathematischer Schulleistungen des jeweils selben Messzeitpunktes untereinander so verbunden, dass die korrelativen Zusammenhänge ermittelt werden können. Das Latent-State-Modell wird mit Berücksichtigung der Korrelationen der nicht beobachtbaren Messfehler und situationsspezifischen Einflüsse mit latenten Variablen, Faktorladungen, Residualvarianzen und Residualkorrelationen spezifiziert. Dieses Design wird jeweils für die Messzeitpunkte eins, zwei, drei und vier erstellt.

Mit einem dann konzipierten Strukturmodell, bestehend aus den beiden Messmodellen der Konstrukte kann die Unabhängigkeit der KFT-Dimensionen verbale Fähigkeiten, quantitative Fähigkeiten und nonverbale Fähigkeiten hinsichtlich der Skalen Sachrechnen, Geometrie und Arithmetik des Mathematikschulleistungstests der DEMAT-Reihe auf latenter Ebene überprüft werden. Es wird angenommen, dass die Konstruktunabhängigkeit nachgewiesen werden kann. Die Mittelwerte der Varianz der latenten Variablen umfassen hier die Messfehler, die immer wieder auftreten. In der untenstehenden Abbildung sind die Basismessmodelle ersichtlich.

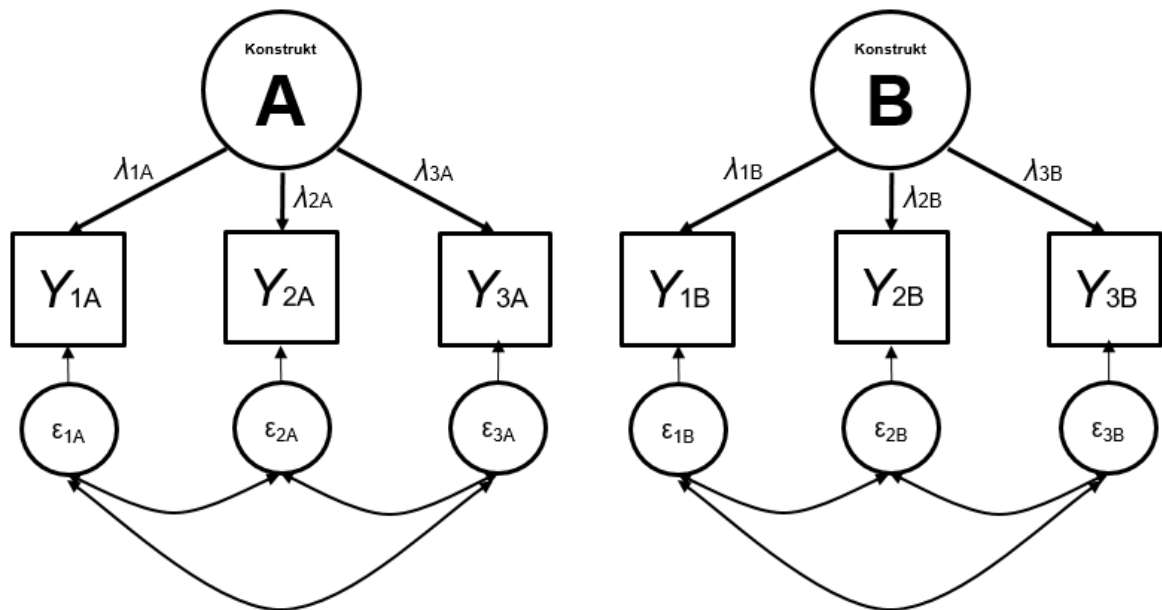


Abbildung 12 Basis-Messmodelle beider Konstrukte;
 λ = Faktorladung; Y = manifeste Variable; ε = Residuum.

Die folgenden Abbildungen zeigen das Analysedesign für die Messverfahren entsprechend der jeweiligen Konzeptionen für die Klassenstufen.

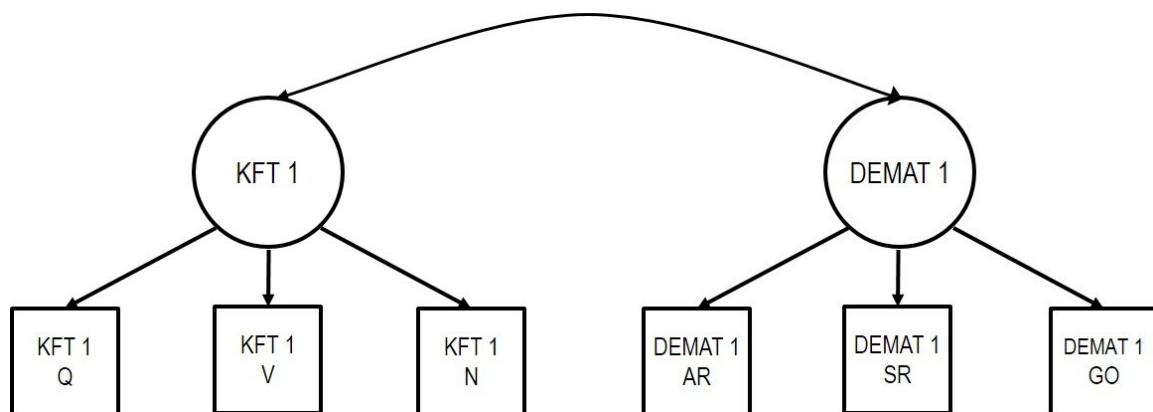


Abbildung 13 Faktorielle Differenziertheit KFT 1 vs. DEMAT 1;
 Q = quantitativ; V = verbal; N = nonverbal; AR = Arithmetik; SR = Sachrechnen;
 GO = Geometrie.

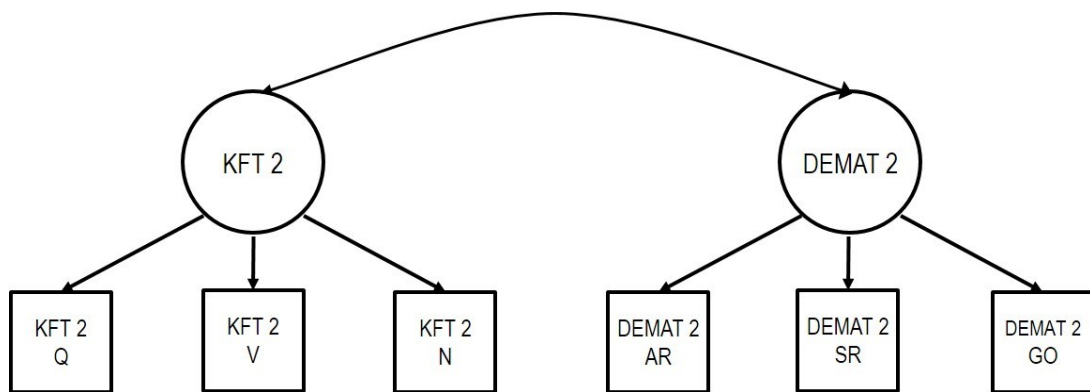


Abbildung 14 Faktorielle Differenziertheit KFT 2 vs. DEMAT 2;
Q = quantitativ; *V* = verbal; *N* = nonverbal; *AR* = Arithmetik; *SR* = Sachrechnen;
GO = Geometrie.

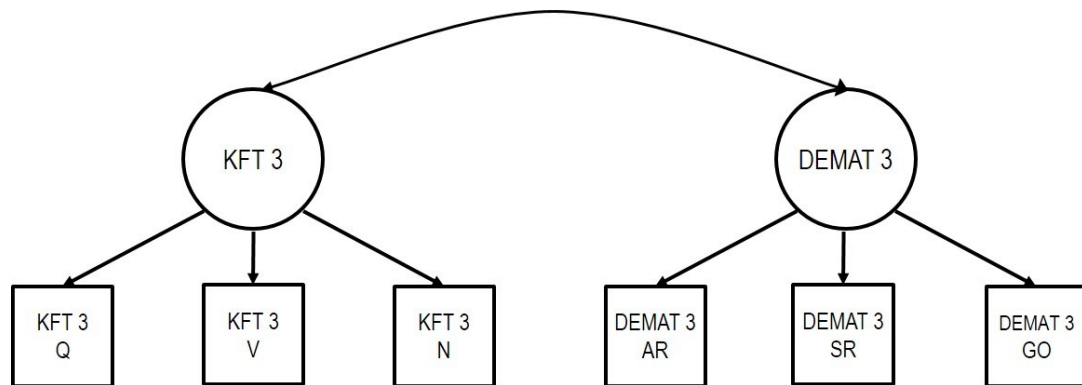


Abbildung 15 Faktorielle Differenziertheit KFT 3 vs. DEMAT 3;
Q = quantitativ; *V* = verbal; *N* = nonverbal; *AR* = Arithmetik; *SR* = Sachrechnen;
GO = Geometrie.

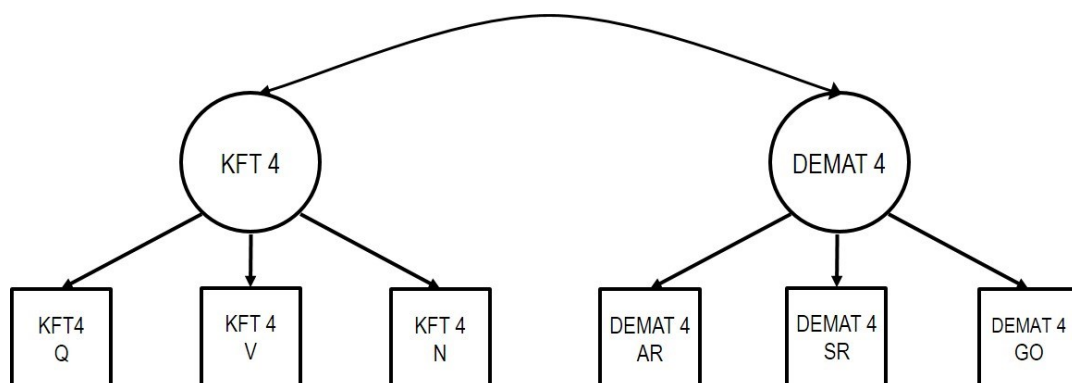


Abbildung 16 Faktorielle Differenziertheit KFT 4 vs. DEMAT 4;
Q = quantitativ; *V* = verbal; *N* = nonverbal; *AR* = Arithmetik; *SR* = Sachrechnen;
GO = Geometrie.

Nachfolgend sollen nun die Ergebnisse berichtet werden.

3.4.2 Analysestrategie 2 – Invarianzprüfung

Das Vorgehen der Analyse der Messinvarianzprüfung in der vorliegenden Arbeit folgt den „Best Practice“-Ansätzen für die Feststellung von Messgleichwertigkeit nach den von Geiser (2011) bzw. nach Widaman und Reise (1997; auch Meredith, 1993 In Geiser, 2011) empfohlenen Strategien. Mehrere Modelle mit zunehmenden Annahmen der Gleichheit (zunehmende Restriktionen) werden getestet.

Ist konfigurale Invarianz vorhanden, so lassen sich die Beziehungen zwischen den latenten Variablen über die Guppen bzw. Messverfahren hinweg vergleichen. Die Faktorenstruktur ist unverändert über die Testverfahren bzw. Gruppen und die Anzahl der Faktoren und das Ladungsmuster sind invariant. Fehlt die konfigurierte Invarianz, so werden unterschiedliche Konstrukte gemessen, die nur schwerlich zu interpretieren sind (Geiser, 2011). Auf einer nächsten Stufe sollten zusätzlich die Faktorladungen gleich sein. Bei diesem Grad der Invarianzprüfung wird geprüft, inwieweit die Verfahren gleich diskriminieren. Bei der faktoriellen (metrischen) Invarianz üben die Items jeder Indikatorengruppe den gleichen Einfluss der latenten Variablen auf die entsprechenden manifesten Variablen aus, angezeigt über die Faktorladung (Geiser, 2011). Ist metrische Invarianz vorhanden, so sind die Beziehungen zwischen den latenten Variablen über die Testverfahren hinweg vergleichbar und die Pfade sind gruppenspezifisch über die Testverfahren identisch (Geiser, 2011). Fehlt die metrische Varianz, so kann die Testkonstruktion konzeptuell unterschiedlich sein oder es sind bestimmte Items eines Tests zu einem Zeitpunkt besser anwendbar als zu einem anderen Zeitpunkt.

Die Messinvarianz erfolgt über Gleichsetzungsrestriktionen der oben genannten Parameter. Diese gehen davon aus, dass Parameter gleichgesetzt werden, sodass sie auf den gleichen Wert geschätzt werden können (Geiser, 2011). Wenn die faktoriellen Invarianzbedingungen zufriedenstellend sind, beruhen Werte latenter Variablen zu jedem Zeitpunkt der Messung auf der gleichen Metrik und spätere Schlussfolgerungen sind gerechtfertigt (Widaman, Ferrer & Conger, 2010).

Die Entscheidung darüber, welche Invarianzen vorliegen, erfolgt über Modellvergleiche. Mehrere Modelle mit zunehmenden Annahmen der Gleichheit (zunehmende Restriktionen) werden getestet. Zunächst werden die die Residualkorrelationen gesichtet, um die Korrelationen zwischen den latenten Variablen zu prüfen. Die Pfade geben den Einfluss

der latenten Variablen auf die Indikatoren (Ladung) an. Manifeste Variablen besitzen Varianz und Intercept (Mittelwerte) ebenso wie das Residuum. Die Kontrolle der Residualkovarianzen ist nötig, weil Messfehlervarianzen auch methodenspezifität beinhalten die über die Zeit korreliert sind. Als Fit-Indizes werden nach Empfehlung von Kline (1998) mehrere Parameter genutzt: Neben den Statistiken des χ^2 -Tests (χ^2 , df, p) soll das Verhältnis von χ^2 und den Freiheitsgraden berichtet werden. Für einen akzeptablen Fit sollte sich ein Wert < 2.5 , besser noch < 3 ergeben (Hair, Anderson, Tatham & Black, 1998 in Geiser, 2011). Besonders in großen Stichproben ist es wahrscheinlich, dass auch kleinere Abweichungen vom Modell in einem signifikanten χ^2 -Test resultieren. Daher sollen weiterhin der Normed-Fit-Index (NFI) und der Comparative-Fit-Index (CFI) berichtet werden. Der Root Mean Square Error of Approximation (RMSEA) sollte für eine adäquate Modellpassung $< .08$ betragen. Die Nullhypothese kann beibehalten werden, wenn die untere Intervallgrenze des 90 % – Konfidenzintervalls kleiner als .05 ist (Rose et al., 2010). Je größer die Diskrepanz zwischen Modell und Populationsebene, desto schlechter die Modellpassung. Die Werte können Werte zwischen 0 und 1 annehmen. Vorteil des RMSEA ist eine geringe Beeinflussung durch die Stichprobengröße. Der Standardized-Root-Mean-Square-Residual-Koeffizient (SRMR) ist ein standardisiertes Instrument für die Beurteilung der Residuen. Der SRMR gibt das mittlere absolute Residuum der Residualkorrelationen an und hat bei Werten zwischen 0 bis $< .05$ eine perfekte bzw. gute Modellpassung. Werte $.05 < \text{SRMR} < .10$ zeigen eine adäquate bis mäßige Modellpassung (Geiser, 2011). Es werden die Stichprobenkennwerte minus modellimplizierte Kennwerte berechnet. Werte $< .05$ sind wünschenswert, sie sagen aus, dass mit Hilfe des Modells die beobachteten Varianzen, Kovarianzen und ggf. Mittelwerte angemessen reproduziert werden (Geiser, 2011).

Anhand des χ^2 -Differenzentests und der Modellfitindices (χ^2 -Test, $\text{CFI} > .95$, $\text{RMSEA} < .08$, $\text{SRMR} < .08$) erfolgt die Entscheidung darüber, welches Modell auf die vorliegende Datenstruktur passt. Können bestimmte Grade der beabsichtigten Invarianz nicht erreicht werden, so empfiehlt Sass (2011), nicht invariante Items aus der Analyse zu entfernen, was allerdings nur dann möglich ist, wenn eine große Anzahl Items zur Verfügung steht. Ein Vergleich mit früheren Untersuchungen oder anderen Eichstichproben ohne Invarianz ist dann kaum möglich.

Als Alternative kann man ein Modell mit partieller Invarianz annehmen, in dem nur für bestimmte Items Vergleiche zwischen den latenten Variablen möglich sind. Empfehlenswert ist diese Methode allerdings nur, wenn die fehlende Invarianz einzelner Items ins-

gesamt nur geringen Genauigkeitsverlust mit sich bringt. Alternativ kann auch bei begründeter Annahme, dass sich die Invarianz nur gering auswirkt, die Analyse unter Einbezug aller Items vollzogen werden, so die Empfehlung von Widaman et al. (2010). Lässt sich eine skalare Invarianz zwischen den Testverfahren der verschiedenen Zeitpunkte nachweisen, so sind auch Mittelwertvergleiche der latenten Variablen möglich. Kann eine strikte Messinvarianz festgestellt werden, so unterscheiden sich Gruppen- bzw. Testverfahren der verschiedenen Zeitpunkte lediglich auf latenter Ebene und die Ladungen, Intercepts und Messfehlervarianzen sind identisch. Bei der metrischen Invarianzprüfung wird geprüft, inwieweit die Verfahren gleich diskriminieren. Die Items jeder Indikatorengruppe üben den gleichen Einfluss der latenten Variablen auf die entsprechenden manifesten Variablen aus, angezeigt über die Faktorladung. Die Akzeptanz von nur partieller faktorieller Invarianz ist dann zulässig, wenn die Untersuchung auf fundierten Überlegungen über die Eigenschaften, die einem Konstrukt zugrunde liegen, basiert und die Operationalisierung entsprechend gründlich erfolgte. Eine vollfaktorielle Invarianz über die Verfahren hinweg wird sich selten erreichen lassen, aber leicht unterschiedliche Eigenschaften der zugrunde liegenden Konstruktion werden als partielle Invarianz toleriert, wenn eine hinreichende Begründung zur Erklärung der Unterschiedlichkeit vorgenommen werden kann (Knight & Zerr, 2010). Wenn sich keine Invarianz in der Datenstruktur nachweisen lässt, sind Gruppenvergleiche oder auch Vergleiche über mehrere Testzeitpunkte mit wechselnden Verfahren nicht möglich und die Messinstrumente sind für den gewünschten Vergleich nicht zu empfehlen bzw. bedürfen einer Revision (Widaman et al., 2010; Knight & Zerr, 2010).

Für die schrittweise Beantwortung der Fragestellung 2 wird erstens ein Messmodell spezifiziert, in dem die Annahme gilt, dass über die vier Jahre die gleichen Konstrukte gemessen werden, sodass eine hohe Vergleichbarkeit der Items des kognitiven Fähigkeitstests für die vier Messzeitpunkte festzustellen ist. Ebenso wird eine hohe Vergleichbarkeit der Mathematikleistungsbereiche untereinander im betrachteten Zeitraum angenommen. Im Messmodell wird spezifiziert, inwieweit latente Variablen (Faktoren) durch beobachtete (manifeste Indikatoren) gemessen werden. Diese werden aus den oben genannten Gründen explizit in einem Latent-State-Modell berücksichtigt. So kann die Zusammenhangsstruktur zwischen den latenten Variablen korrelativ beschrieben werden.

Die Verknüpfung der Indikatoren mit ihren Subskalen und den zugehörigen latenten Variablen folgt dem Prinzip von Strukturgleichungsmodellen. Zudem sind im Strukturmodell manifeste Variablen enthalten, die mit den latenten Variablen über Pfade oder Korrelationen zusammenhängen und dadurch fehlerfrei gemessen werden können. Die Residuen

sind untereinander mit den manifesten Indikatoren korreliert. Die Parameter eines einfachen Messmodells sind in der folgenden Abbildung veranschaulicht.

Das im ersten Schritt zu spezifizierende Messmodell enthält zunächst die manifesten Variablen, die von den latenten Variablen beeinflusst werden. Die Höhe des Einflusses auf die manifesten Variablen wird über die Faktorladung ermittelt. Zudem wirkt auf jede manifeste Variable ein Residuum, wie die folgende Abbildung zeigt.

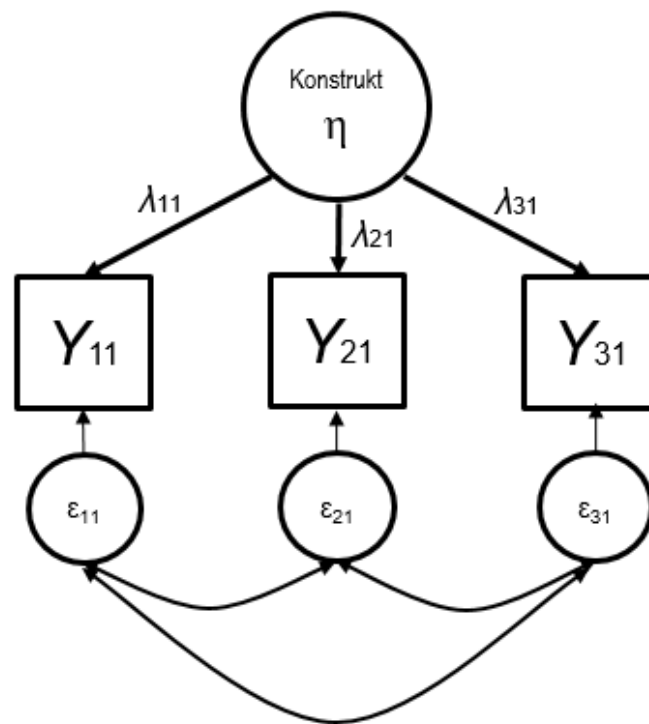


Abbildung 17 Latent-State-Model – Messmodell
 η = latente Variable; λ = Faktorladung; Y = manifeste Variable; ε = Residuum.

Die Pfade geben den Einfluss der latenten Variablen auf die Indikatoren an (Faktorladung). Die manifeste Variable enthält einen Mittelwert (Intercept) und eine Varianz. Das Residuum des nicht erklärbaren Anteils besitzt eine Varianz, die auf die manifesten Variablen wirkt. Die auf die Datenstruktur am besten passenden Messmodelle werden zu einem Strukturgleichungsmodell verknüpft.

Im unten dargestellten anschließend spezifizierten Strukturmodell wird die Zusammenhangsstruktur zwischen den latenten Variablen korrelativ mit dem Ziel beschrieben, Hypothesen über den latenten Beeinflussungsgrad berücksichtigende Pfadanalyse abzuschätzen.

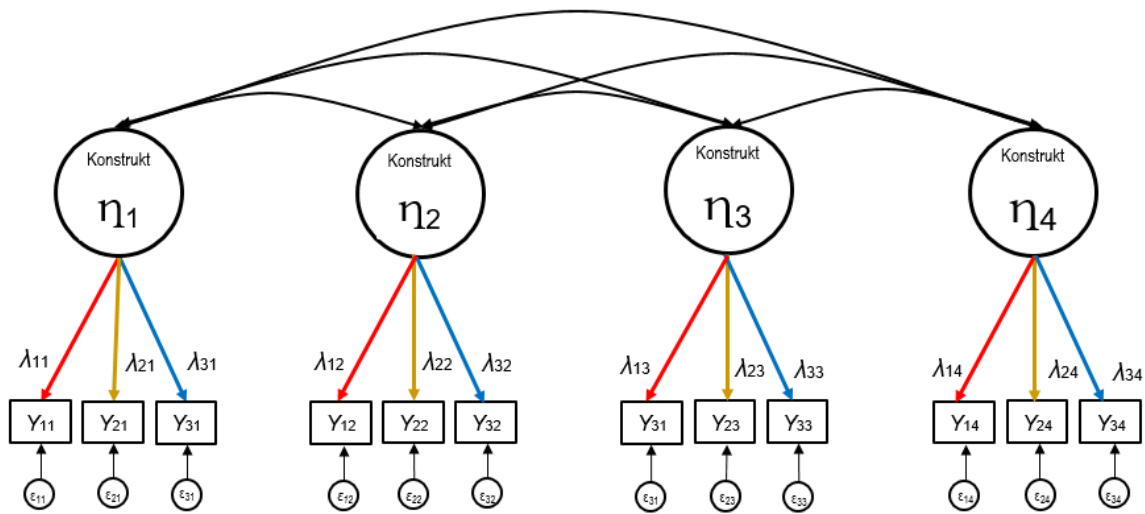


Abbildung 18 Latent-State-Model I – Strukturmodell

η = latente Variable; λ = Faktorladung; Y = manifeste Variable; ϵ = Residuum.

Die obige Abbildung zeigt in einem ersten Schritt die Verknüpfung der Messmodelle und in der folgenden Abbildung die Spezifikation des Strukturmodells. Es wird jeweils ein Modell für die Testverfahren der DEMAT-Reihe und für die Testverfahren der KFT-Reihe modelliert.

Für die Testung der Messinvarianz der Testverfahren der kognitiven Fähigkeiten wurden jeweils die Items der verbalen, quantitativen und nonverbalen Subtests des KFT zur Invarianzprüfung über die vier Messzeitpunkte (Daten der Kinder aus der ersten, zweiten, dritten und vierten Klassenstufe) herangezogen, wie die folgende Abbildung zeigt.

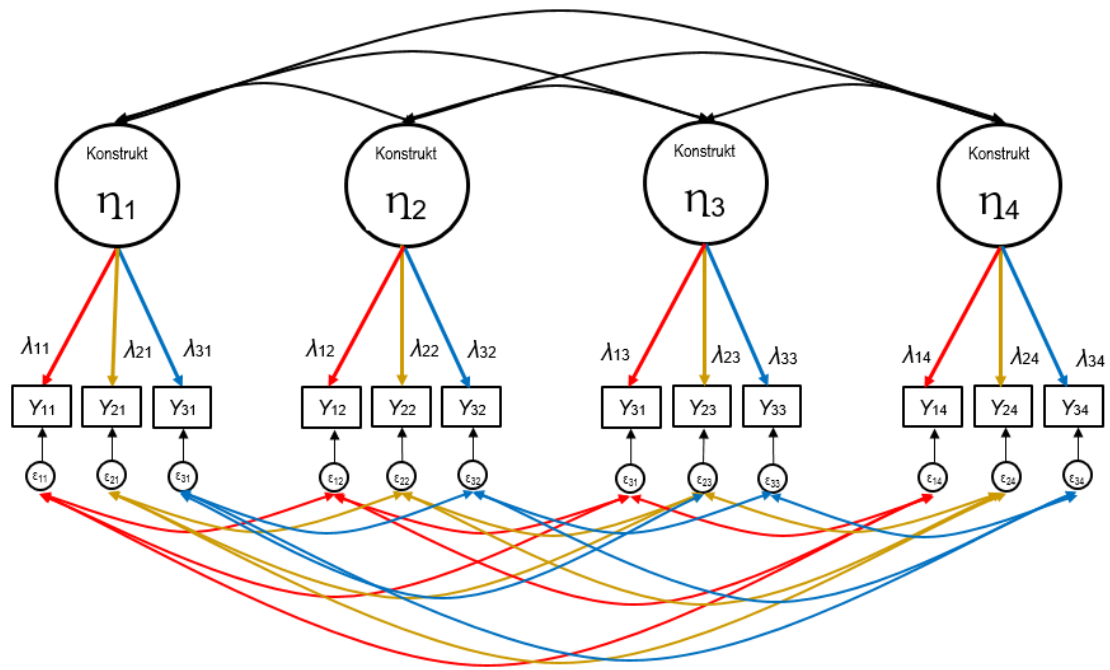


Abbildung 19 Latent-State-Model II – Strukturmodell

η = latente Variable; λ = Faktorladung; Y = manifeste Variable; ε = Residuum.

Das Ausgangsmodell bildeten Faktoren der jeweiligen Subskala, wobei die Faktoren der ersten Ordnung auf zwei zueinander in Beziehung stehenden latenten Faktoren zweiter Ordnung regredierte werden. Zum ersten Messzeitpunkt (Klassentufe 1) werden 72 Items aus den drei Subtests, zum zweiten Messzeitpunkt 72 (Klassenstufe 2), 140 Items des dritten Messzeitpunktes und 195 Items zum Messzeitpunkt 4 analysiert.

Für die Invarianzprüfung der Deutschen Mathematiktestreihe DEMAT 1 mit 36 Items, DEMAT 2 (57 Items), DEMAT 3 (37 Items) und DEMAT 4 mit 40 Items werden latente Variablen an die manifesten Variablen gebunden und deren Einflusshöhe und Korrelation ermittelt (siehe Abbildung oben). Zunächst wird eine schwache Restriktion vollzogen, welche die Gleichheitsrestriktion auf die Faktorladungsstruktur (Anzahl der Faktoren und das Ladungsmuster) vorsieht. Der Subtest Arithmetik wird fixiert, da diesem der größte Erklärungswert der mathematischen Kompetenzen zukommt. Die Arithmetikanteile besaßen zudem die beste Itemstruktur über die Zeit und lieferten die höchste Anzahl an Items. Die Schätzung des zweiten und dritten Indikators erfolgt frei. Ebenso werden die Erwartungswerte der latenten Variablen frei geschätzt. Im folgenden Abschnitt soll die Analysestrategie zur Prüfung der autoregressiven Effekte vorgestellt werden.

3.4.3 Analysestrategie 3 – Autoregressive Effekte

Die vorgestellten Modelle sind geeignet, die interindividuelle Stabilität zu ermitteln, nicht aber um intraindividuelle Niveauveränderungen in den erfassten Leistungen festzustellen. Die Analyse der Frage nach der interindividuellen Stabilität erfolgt auf Basis des zuvor spezifizierten Messfehlers berücksichtigenden Latent-State-Modells mit konfigurabler Invarianz, welches sich bei der Modelltestung zur Invarianzprüfung als besonders geeignet erwiesen hat.

Es besteht aus einem Messmodell und dem Latent-State-Modell, welches bereits in der Analyse der Frage 2 schrittweise entwickelt und getestet wurde. Dieses wird zu einem Strukturmodell spezifiziert, das autoregressive Beziehungen über vier Messzeitpunkte berücksichtigt. In diesem wird die Beschreibung der Zusammenhangsstruktur latenter Variablen auf korrelativer Ebene erfolgen. Weiter werden die regressive Beziehung der Faktorparameter, die Mittelwertstrukturen sowie die Messfehlervarianzen aufgeklärt.

Die Analyse der Stabilität beider Konstrukte erfolgt unter Berücksichtigung latenter Einflüsse und folgt einem schrittweisen Modellaufbau. Dabei werden in einem ersten Schritt die autoregressiven Effekte beider Konstrukte (kognitive Fähigkeiten und Mathematikleistungen) getrennt ermittelt. Dafür werden also zunächst die Einflussfaktoren aus dem jeweils vorherigen Messzeitpunkt auf den folgenden Messzeitpunkt berechnet, wie in der folgenden Abbildung gezeigt.

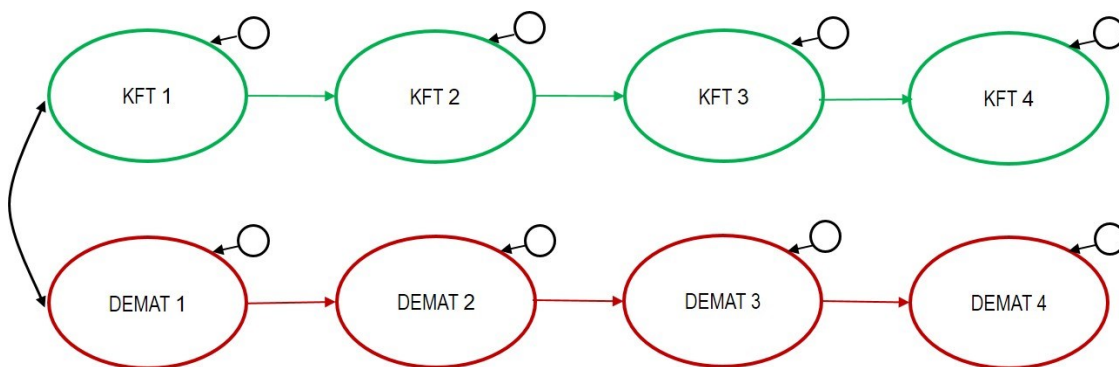


Abbildung 20 Autoregressives Modell 1. Ordnung.

Die Stabilität über die Zeit wird über ein autoregressives Modell (ARM) separat für kognitive Fähigkeiten, gemessen mit dem KFT der jeweiligen Messzeitpunkte und dem DEMAT zu den vier Messzeitpunkten und separat für die mathematischen Schulleistungen ermittelt.

Der grüne Bereich stellt die kognitiven Fähigkeiten (KFT) dar. Für die mathematischen Schulleistungen (DEMAT), die hier rot gezeichnet sind, erfolgt die Zulassung der Regressionen auf die gleiche Weise. Auf die zeichnerische Darstellung aller berücksichtigten Parameter wird aus Gründen der Übersichtlichkeit verzichtet. Die komplette Darstellung befindet sich auf der beigelegten CD im Anhang.

Im Anschluss erfolgt die Anbindung des ersten Messzeitpunktes auf den übernächsten, also dritten Messzeitpunkt und die Anbindung des zweiten Messzeitpunktes mit seinen Beeinflussungsfaktoren auf den vierten Messzeitpunkt, um zu ermitteln, inwieweit Ergebnisse von Messungen auf früheren Messungen basieren und damit stabil sind. Die untenstehende Abbildung veranschaulicht das Vorgehen. In der Modellspezifikation gilt die Annahme, dass Regressionen über die vier Messzeitpunkte bestehen.

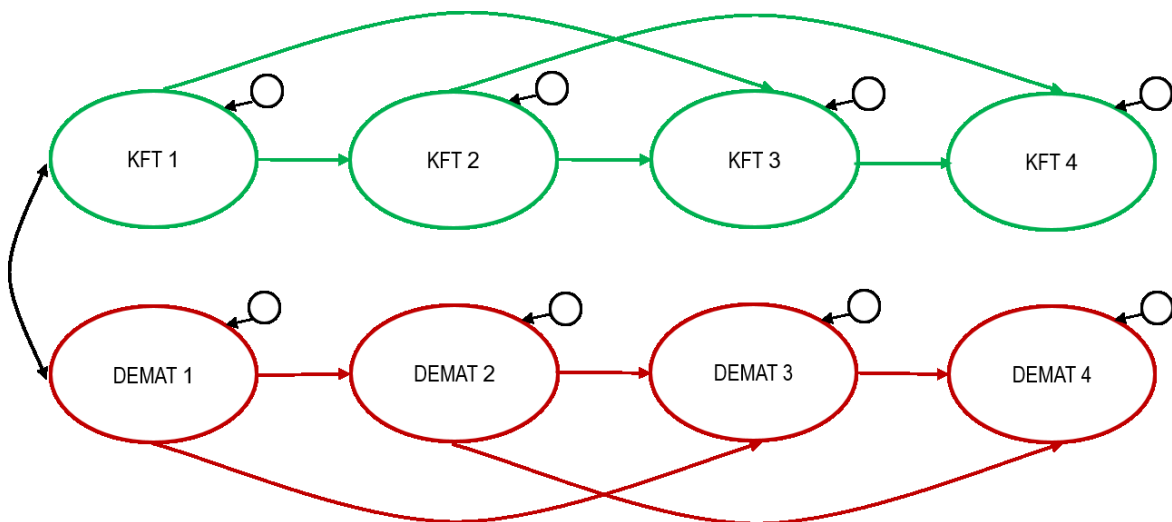


Abbildung 21 Autoregressives Modell 2. Ordnung.

Die Pfade verlaufen zunächst direkt vom vorherigen Zeitpunkt auf den Folgenden und geben damit die autoregressiven Pfade an. Dann folgt eine Hinzunahme der Pfade von KFT 1 auf den KFT 3, also ein Modell der 2. Ordnung.

Für die Aufklärung der Stabilitätsbeziehungen über zwei Messzeitpunkte werden nun die regressiven Pfade in dem autoregressiven Modell (ARM) separat für kognitive Fähigkeiten und mathematischen Schulleistungen von einem früheren Messzeitpunkt auf den übernächsten zugelassen. Auf die zeichnerische Darstellung aller berücksichtigten Parameter wird hier ebenfalls in der gezeigten Grafik aus Gründen der Übersichtlichkeit verzichtet. Die komplette Darstellung befindet sich auf der beigelegten CD im Anhang.

Dann erfolgt die Anbindung über 3 Messzeitpunkte, womit die regressive Beziehung einer 3. Ordnung ermittelt werden kann, wie die untenstehende Abbildung über die regressionsverbindungen vom ersten auf den vierten Messzeitpunkt veranschaulicht.

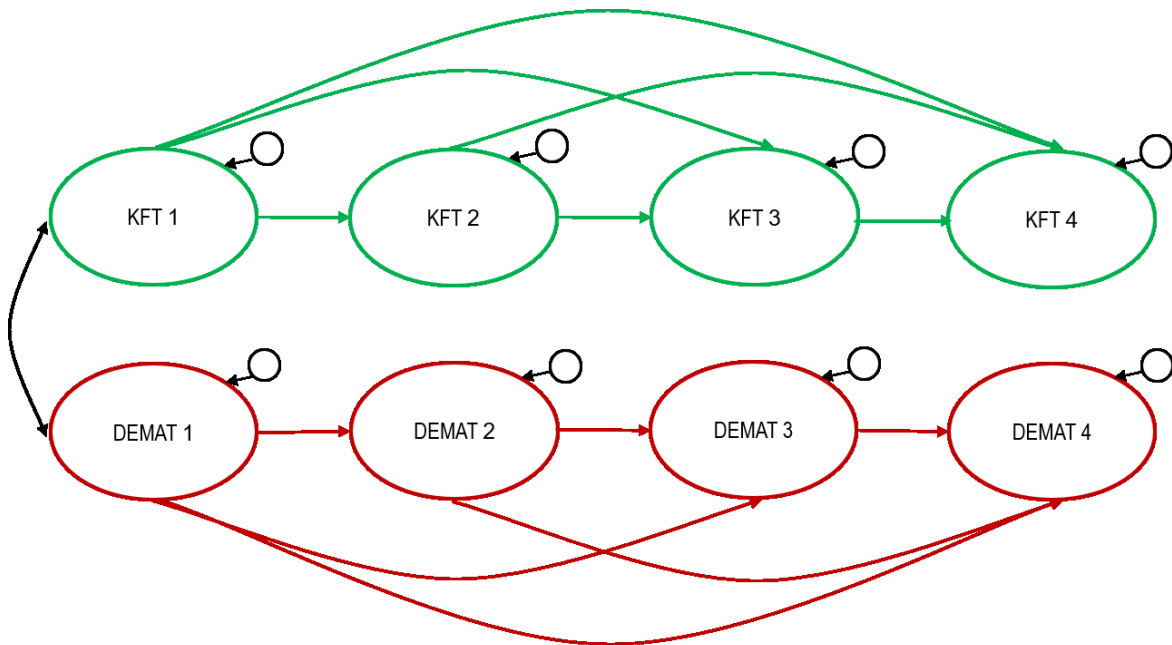


Abbildung 22 Autoregressives Modell 3. Ordnung.

Um auch die Stabilität über die gesamte Grundschulzeit der Kinder zu ermitteln, werden nunmehr die autoregressiven Effekte der dritten Ordnung spezifiziert, indem die Regressionen sowohl für kognitive Fähigkeiten und mathematische Schulleistungen vom 1. auf den 3. Messzeitpunkt zugelassen werden. Der oberste Pfeil weist die regressive Beziehung des 1. Messzeitpunkt auf den 4. Messzeitpunkt aus. Die Darstellung aller berücksichtigten Parameter befindet sich aus Gründen der Übersichtlichkeit auf der beigelegten CD im Anhang.

3.4.4 Analysestrategie 4 – Kreuzverzögerte Effekte

Als Indikatoren für kognitive Fähigkeiten gelten verbale, quantitative und nonverbale Itemgruppen. Für die Bewertung mathematischer Fähigkeiten wurden Items aus den Bereichen Geometrie, Arithmetik und Sachrechnen bewertet und zunächst in einem autoregressiven Modell spezifiziert, welches die Einflusspfade von Messergebnissen auf ihre Prädiktionskraft hinsichtlich der Ergebnisse zu einem späteren Zeitpunkt untersucht und in einem zweiten Schritt wurden Crossed-Lagged-Effekte untersucht. Um die Einflussrichtung der kogni-

tiven Fähigkeiten auf die mathematischen Schulleistungen zu bestimmen, sind Strukturgleichungsmodelle geeignet. Zunächst wird der gerichtete Zusammenhang zwischen den Items der jeweiligen Subskalen des kognitiven Fähigkeitstests des ersten Messzeitpunktes als unabhängige Variablen auf die Subskalen des Mathematikleistungstests des zweiten Messzeitpunktes als abhängige Variablen verbunden. Es erfolgt weiter eine Spezifikation der Prädiktorvariablen der kognitiven Subskalen auf den dritten Messzeitpunkt der mathematischen Schulleistungen und analog die Spezifikation der Prädiktorvariablen der kognitiven Fähigkeitsskalen auf die Subskalen der mathematischen Leistungen zum vierten Messzeitpunkt. Die Annahme multipler theoretischer Zusammenhangsstrukturen über Messfehler, Kovarianzen und Regressionen von mehreren Variablen ist sehr komplex und der Modelaufbau erfolgt zur Vermeidung von Syntaxfehlern entsprechend den Empfehlungen von Geiser (2011) schrittweise. Das zuvor spezifizierte autoregressive Modell 3 bildet dabei die Grundlage.

Zusätzlich zu dem einfachen autoregressiven Modell 3 soll hier jedoch der Einfluss kognitiver Fähigkeiten auf mathematische Schulleistungen des Folgezeitpunktes aufgeklärt werden, dargestellt über die grünen Pfeile in der Mitte der untenstehenden Abbildung.

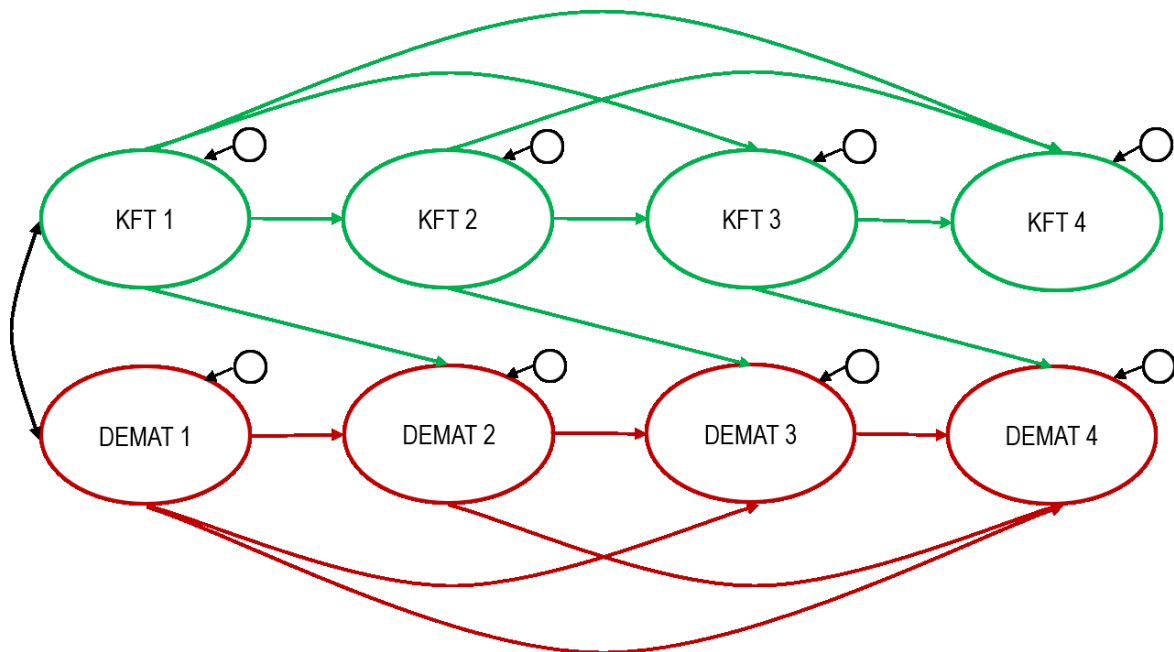


Abbildung 23 Crossed-Lagged-Modell 4. Ordnung.

Zunächst werden die Faktorladungen des ersten Indikators KFT 1-Gesamtwertes auf die verbalen, quantitativen und nonverbalen Subskalen spezifiziert. Die Fixierung der Faktorladungen des zweiten Messzeitpunktes auf die zugehörigen Subskalen erfolgt analog sowohl für die kognitiven Fähigkeiten, als auch für die mathematischen Schulleistungen. Im Anschluss erfolgt die Berücksichtigung der in den autoregressiven Modellen bereits spezifizierten autoregressiven Effekte des jeweils selben Konstrukts auf die Variablen des Konstrukts zum jeweils folgenden Messzeitpunkt und auf dasselbe Konstrukt des übernächsten und darauffolgenden Konstrukts. Auch die Residualvarianzen der Indikatoren sollen geschätzt werden. Es soll damit beantwortet werden, ob vor allem kognitive Leistungen prädiktiv für mathematische Schulleistungen sind.

Das oben abgebildete Modell 4. Ordnung beschreibt zunächst die Einflüsse kognitiver Fähigkeiten auf die mathematischen Schulleistungen. Mit diesem in der obigen Grafik dargestellten Modelldesign lassen sich zum einen autoregressive Einflüsse des gleichen Konstrukts berücksichtigen und zum anderen können regressive Einflüsse kognitiver Fähigkeiten auf spätere mathematische Schulleistungen berücksichtigt werden. Diese Modellkonstruktion berücksichtigt darüber hinaus nur solche Items, deren Vergleichbarkeit über die Zeit sich in Fragestellung 2 nachweisen lässt. In dem Modell der 5. Ordnung (Abbildung unten) wird nunmehr das Modell mit kreuzverzögerten Effekten modelliert.

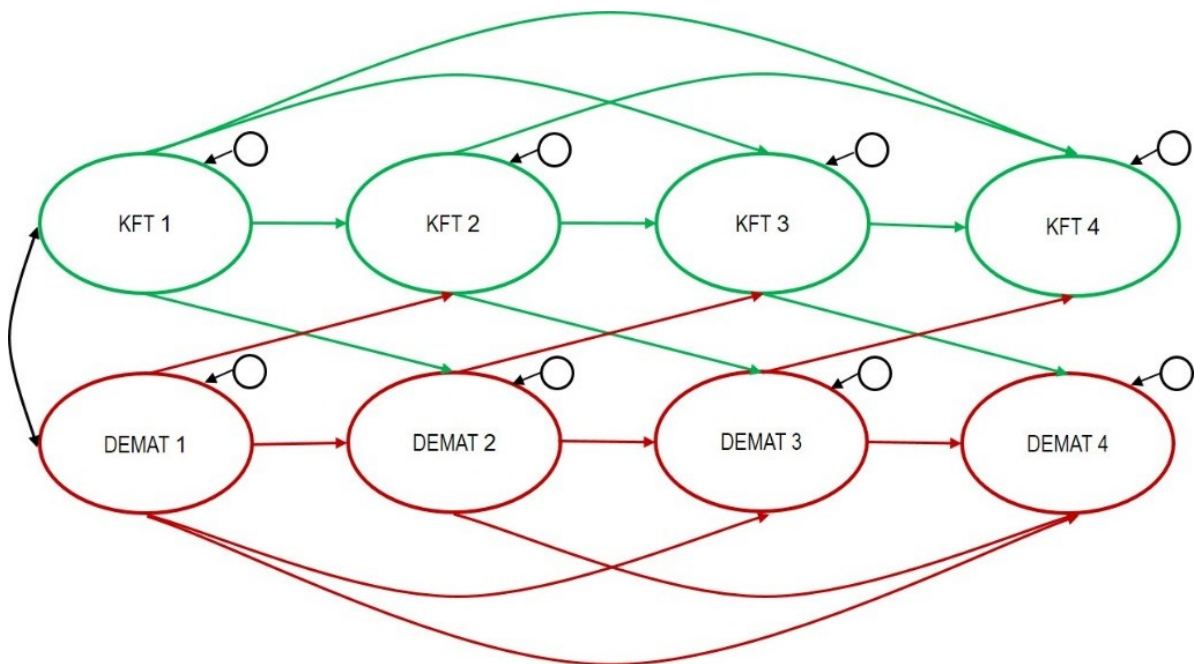


Abbildung 24 Crossed-Lagged-Modell 5. Ordnung.

Für die Analyse werden nun die Regressionspfade der Konstrukte kognitive Fähigkeiten und Mathematikleistungen in einem Crossed Lagged Modell verbunden, sodass Einflüsse des vorherigen Messzeitpunktes des jeweils anderen Konstrukts erfasst werden können. Für die Bestimmung der gegenseitigen Einflusspfade sollen nun auch die Regressionen der mathematischen Schulleistungen als unabhängige Variablen auf die kognitiven Fähigkeitsvariablen des jeweils folgenden Zeitpunktes bestimmt werden. Das abgebildete Modell 5. Ordnung ist das nun angestrebte Modell, das beidseitige Regressionen der vorherigen Messungen mit Wirkungen auf Folgeleistungen ausgeben soll. Diese werden auch kreuzverzögerte Effekte genannt. Sie werden nunmehr auch durch die roten Pfeile in der Mitte, ausgehend vom DEMAT 1 auf den KFT 2, und dann analog vom DEMAT 2 auf den KFT 3 und anschließend vom DEMAT 3 auf die kognitiven Fähigkeiten des 4. Messzeitpunktes.

Jedoch werden entsprechend des empfohlenen schrittweisen Vorgehens des Modelllaufbaus (Geiser, 2011) noch keine Residualkorrelationen zugelassen. Erst in einem weiteren Schritt werden diese nun zugelassen, wie in der folgenden Abbildung, welche das Modell 6. Ordnung dargestellt, gezeigt.

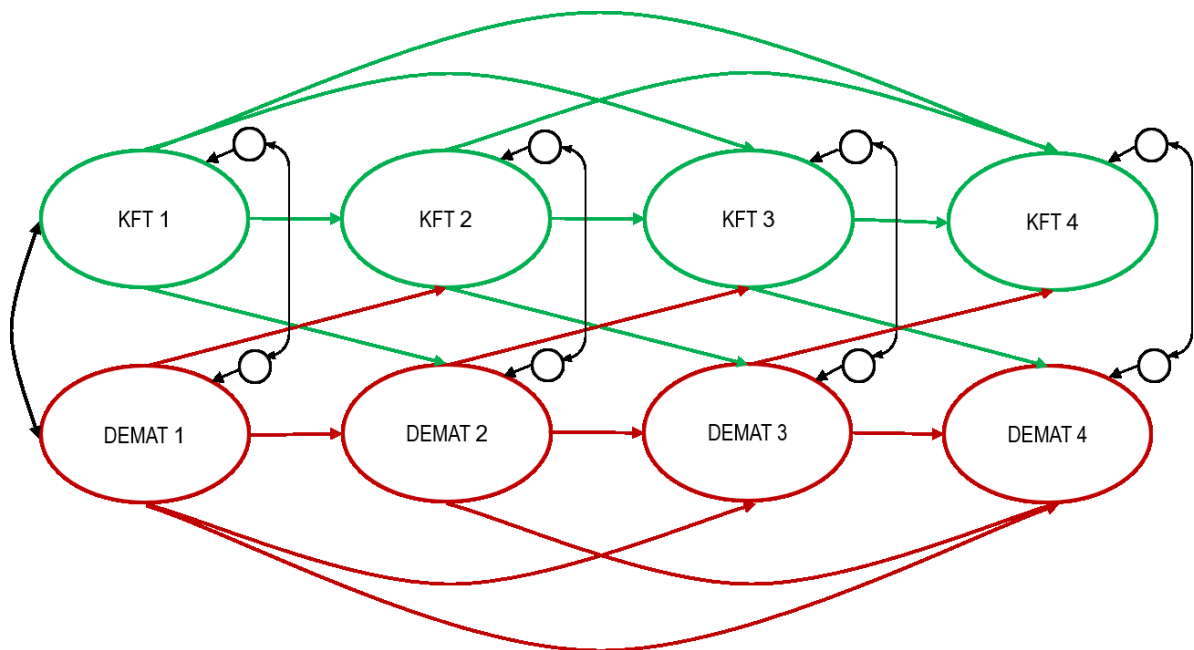


Abbildung 25 Crossed-Lagged-Modell 6. Ordnung.

Diese Spezifikation folgt den im theoretischen Teil der Arbeit ausführlich dargestellten vorherigen Studien, die die Vorwissensaspekte, die in Bildungsinstitutionen erworben werden,

als bedeutsam für die kognitiven Fähigkeiten zu späteren Zeitpunkten berichten (Weber, 2007; Krajewski et al., 2009). Der entscheidende Vorteil der in dieser Arbeit vorgenommenen Spezifikation ist die Berücksichtigung von vier Messzeitpunkten im Grundschulalter sowie die Tatsache, dass die Konstrukte jeweils mit standardisierten Messverfahren erhoben worden sind, die bereichsspezifische Fähigkeiten erfassen und für jede Subskala zahlreiche Items zur Verfügung standen. Das längsschnittliche Design kann daher exaktere Interpretationen erlauben, da die Modellierung mittels latenter Strukturgleichungsmodelle dies zulässt. Trautwein und Köller (2003) sowie Weber (2007) empfehlen ein solches Design, da bisherige Daten vorwiegend querschnittlichen Ursprungs sind.

4 Ergebnisse

Das vierte Kapitel widmet sich der Berichterstattung der Ergebnisse. Zunächst erfolgen die deskriptiven Angaben für den kognitiven Fähigkeitstest und die eingesetzten Mathematikleistungstestverfahren für die vier Messzeitpunkte mit den Angaben zur Normalverteilungsprüfung.

Im Anschluss erfolgt die Ergebnisdarstellung entsprechend der Reihenfolge der Fragestellungen.

Deskriptive Analysen und Normalverteilungsprüfungen

Aus der Gesamtstichprobe konnten von 1726 Kindern zum ersten Messzeitpunkt die Daten von 1560 Kindern zur Auswertung der kognitiven Testverfahren herangezogen werden. Für den zweiten Messzeitpunkt lagen die Daten von 1377 Kindern zugrunde, während zum dritten Messzeitpunkt 1342 und zum vierten Messzeitpunkt 1214 Fälle analysiert werden konnten. Die zur Operationalisierung kognitiver Fähigkeiten eingesetzten Skalen rekrutierten sich aus den jeweiligen Subskalen KFT „verbal“, „quantitativ“ und „nonverbal“ für die jeweiligen vier Klassenstufen.

Die Tabelle zeigt die Fälle zum jeweiligen Messzeitpunkt. Ferner enthält sie die Angaben zu den erreichten minimalen und maximalen Rohwerten in den jeweiligen Subtests, Mittelwerte, Standardabweichungen und Angaben zu Schiefe und Kurtosis der Normalverteilungsprüfung sowie deren Gesamtwerte. Zunächst werden die Angaben für die Messung kognitiver Fähigkeiten in der folgenden Tabelle dargestellt. Für den KFT 1 ist lediglich die Skala des nonverbalen Subtests geringfügig linksschief (.54), aber der Standardfehler dieser Skala rechtfertigt mit .23 dennoch die Annahme einer annähernden Normalverteilung. Für den KFT 2 sind ebenfalls nur geringe Abweichungen erkennbar. Die größte Abweichung

zeigt hier die quantitative Skala mit einer Schiefe von $-.78$, während die Gesamtrohwerte der Normalverteilung entsprechen.

Tabelle 7

Kognitive Fähigkeiten (KFT) für die vier Messzeitpunkte

	<i>N</i>	<i>Min</i>	<i>Max</i>	<i>M</i>	<i>SD</i>	σ	<i>Schiefe</i>	<i>Kurtosis</i>
Messzeitpunkt 1								
KFT 1 verbal	1560	4	43	26.41	6.67	.17	-.24	-.24
KFT 1 quantitativ	1560	0	46	23.27	7.84	.20	.25	-.37
KFT 1 nonverbal	1560	0	45	18.03	8.89	.23	.54	-.14
KFT 1 gesamt		4	122	67.72	19.85	.50	.22	-.39
Messzeitpunkt 2								
KFT 2 verbal	1377	8	46	33.55	6.56	.18	-.72	.10
KFT 2 quantitativ	1377	7	48	34.86	8.33	.22	-.78	-.07
KFT 2 nonverbal	1377	0	47	28.02	9.91	.27	-.34	-.71
KFT 2 gesamt		28	139	96.43	21.93	.59	-.54	-.36
Messzeitpunkt 3								
KFT 3 verbal	1342	0	41	21.03	8.06	.27	-.19	-.35
KFT 3 quantitativ	1342	0	44	28.28	8.15	.27	-.75	.54
KFT 3 nonverbal	1342	0	50	35.27	10.50	.36	-.76	-.11
KFT 3 gesamt		0	130	84.58	23.30	.78	-.62	.27
Messzeitpunkt 4								
KFT 4 verbal	1214	0	68	39.70	12.98	.37	-.35	-.43
KFT 4 quantitativ	1214	0	58	36.41	9.77	.28	-.63	.28
KFT 4 nonverbal	1214	0	70	43.19	13.75	.40	-.68	-.45
KFT 4 gesamt		0	194	119.33	32.28	.93	-.60	-.06

N = Stichprobengröße; *Min* = Minimum; *Max* = Maximum; *M* = Mittelwert; *SD* = Standardabweichung; σ = Standardfehler.

Für die Rohwerte des KFT 3 sind die Voraussetzungen ebenfalls erfüllt, lediglich die non-verbale Skala weist mit einer Schiefe von $-.76$ eine leichte Rechtsschiefe auf. Für die Skalen des KFT 4 kann die Normalverteilung als erfüllt betrachtet werden, wenn auch die Schiefe mit einem Wert von $-.76$ ausgegeben wird, so kann aufgrund des Standardfehlers $.39$ dennoch die Normalverteilung als annähernd gegeben vorausgesetzt werden.

Die Normalverteilungsprüfung erfolgte mit dem Kolmogorov-Smirnov-Test, wie in der nachfolgenden Tabelle dargestellt. Die Ergebnisse des Kolmogorov-Smirnov-Tests zeigen eine dennoch gute Robustheit für die eingesetzten Analyseverfahren. Es ist zudem entscheidend, dass das Merkmal normalverteilt ist (Heller & Perleth, 2000).

Tabelle 8

Tests auf Normalverteilung nach Kolmogorov-Smirnov mit Signifikanzprüfung

	KFT 1	KFT 2	KFT 3	KFT 4
<i>verbal</i>	.06	.11	.05	.06
<i>quantitativ</i>	.06	.12	.08	.08
<i>nonverbal</i>	.09	.08	.13	.12
<i>gesamt</i>	.05	.08	.07	.06

Signifikanz für alle Tests $p = 0.00$

Die nachfolgenden Abbildungen veranschaulichen die Prüfung der Normalverteilung für die kognitiven Fähigkeitstests der jeweiligen Messzeitpunkte grafisch, ausgegeben für die Subtests und den Gesamtwert der zum jeweiligen Messzeitpunkt eingesetzten Testverfahren.

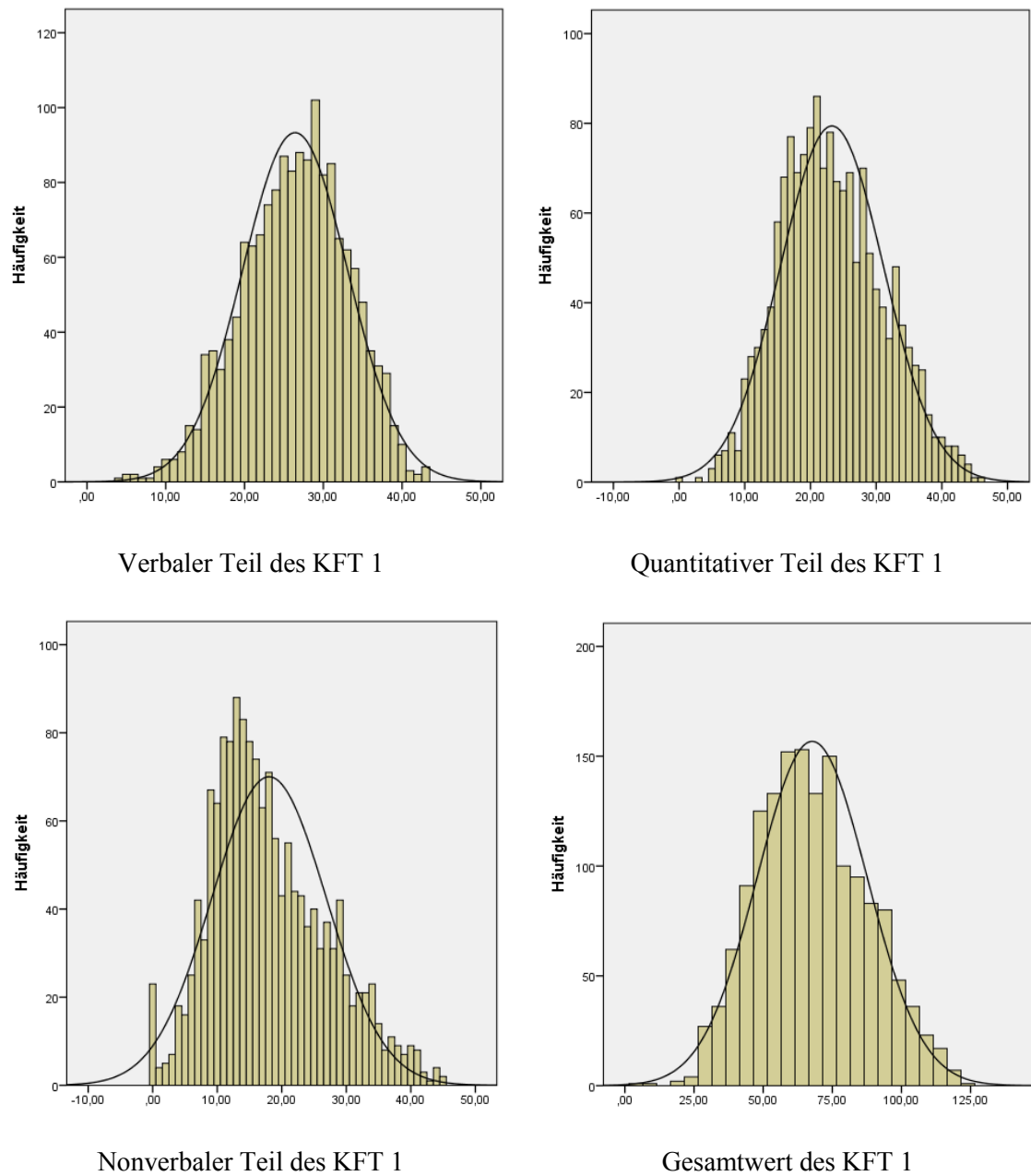


Abbildung 26 Verteilung der Rohwerte zum ersten Messzeitpunkt im KFT

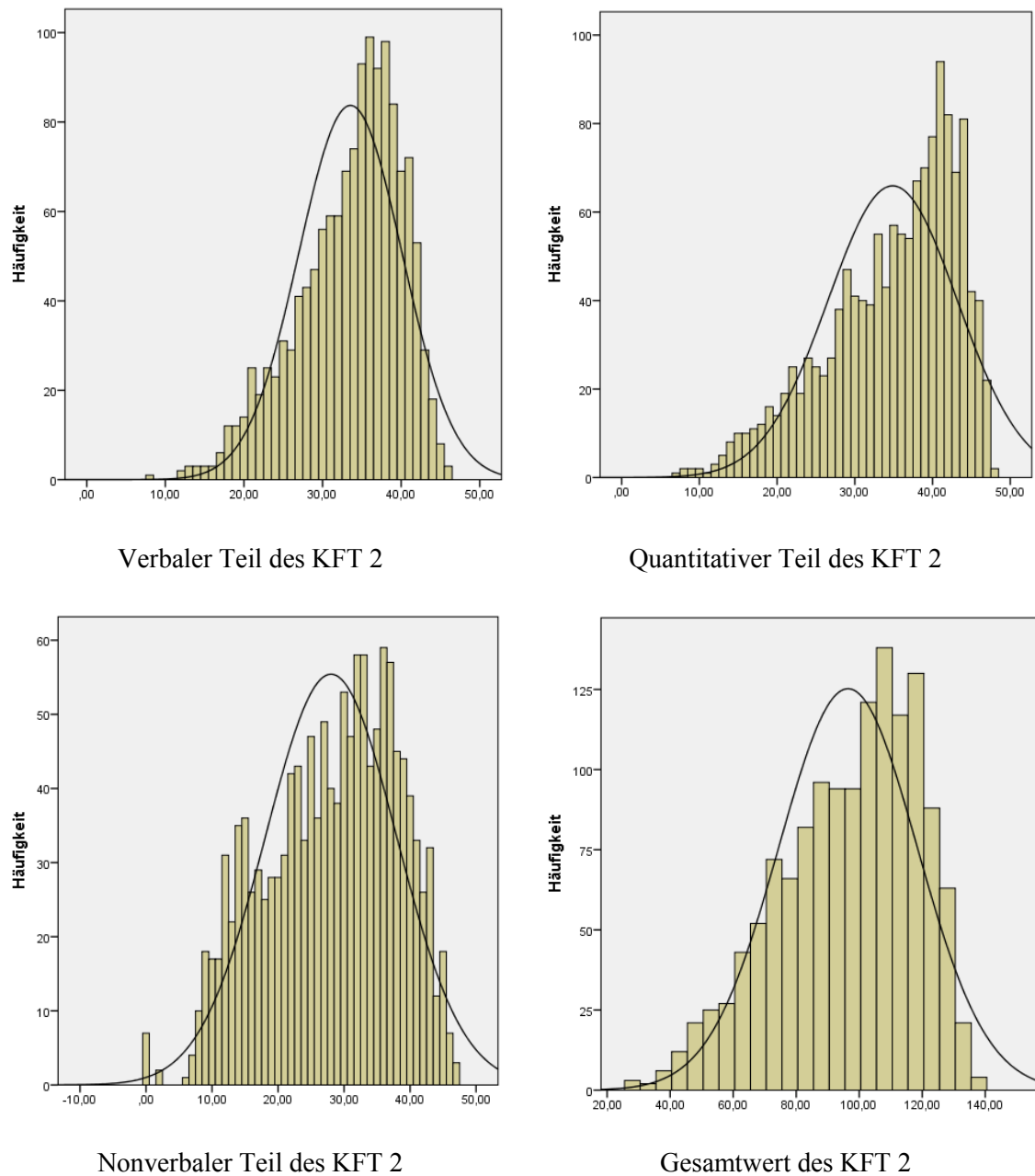


Abbildung 27 Verteilung der Rohwerte zum zweiten Messzeitpunkt im KFT

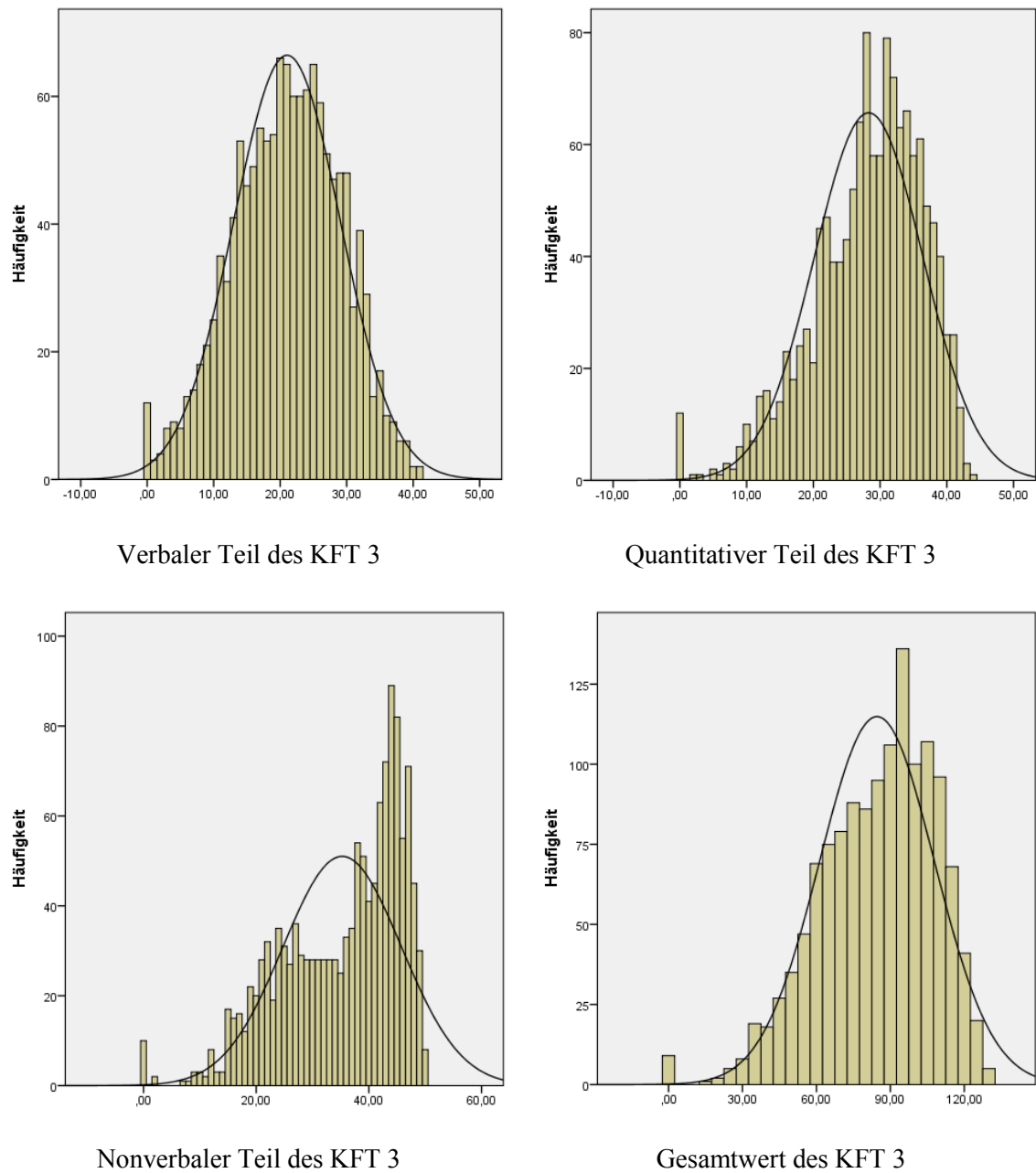


Abbildung 28 Verteilung der Rohwerte zum dritten Messzeitpunkt im KFT

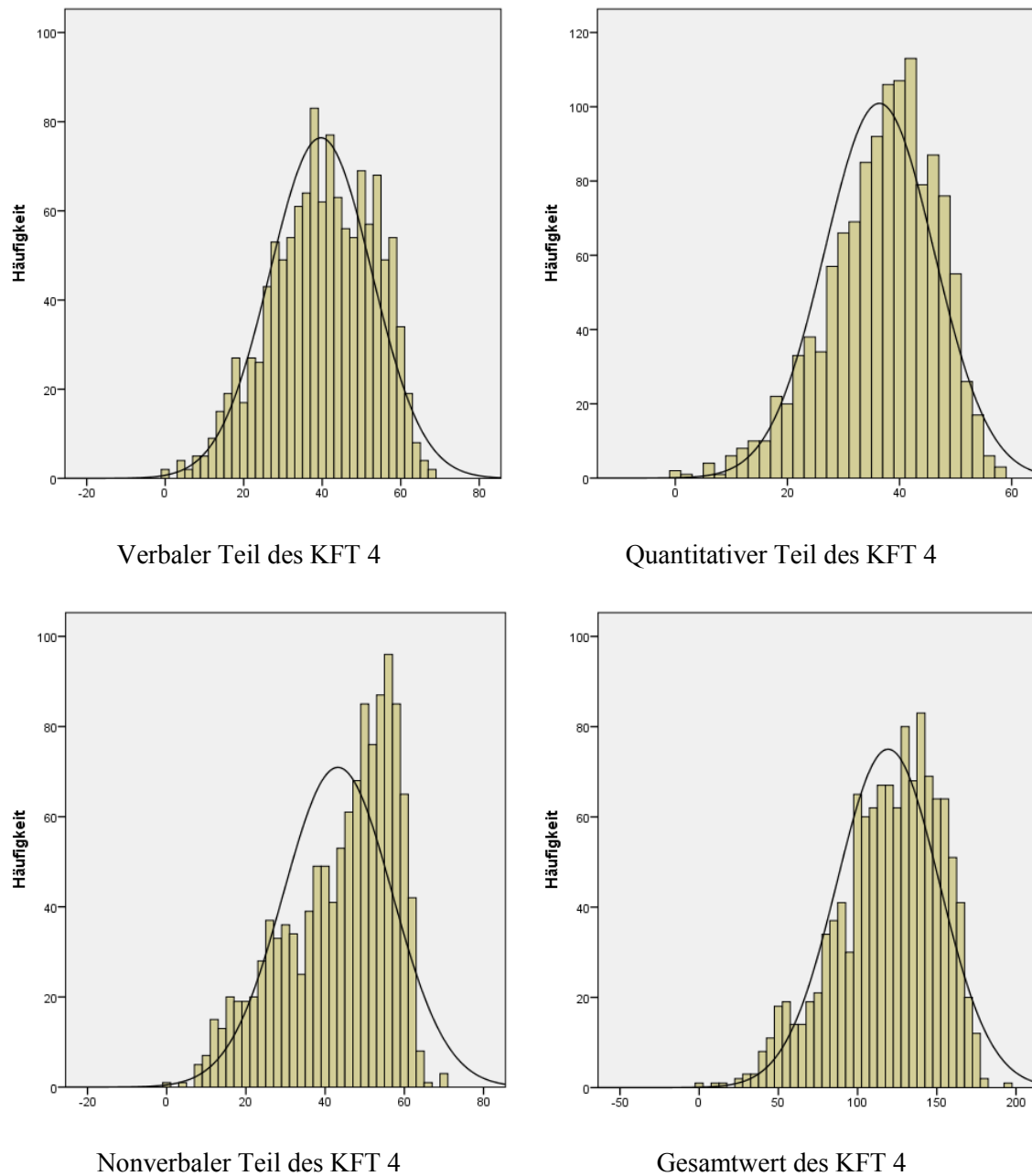


Abbildung 29 Verteilung der Rohwerte zum vierten Messzeitpunkt im KFT

Die Merkmalsausprägung der KFT-Testverfahren wiesen geringe Abweichungen von der Normalverteilung auf und somit hätten den KFT-Werten der Teilnehmer (flächentransformierte) T-Werte zugewiesen werden können, die auf der Normierungstichprobe für die Münchner Hochbegabungstestbatterie (MHBT) aus dem Jahr 1987 beruhen (Heller & Perleth, 2000). Für die Flächentransformation ist das Verfahren nach McCall (1984) mit der

SPSS-RANK-Prozedur geeignet. Da die Daten der bislang noch unveröffentlichten Verfahren des KFT 1, 2 und 3 in die Normierungsstichprobe eingehen werden, wurde von dieser Transformation abgesehen.

Im Anschluss folgen die Angaben zur Normalverteilungsprüfung für die Testverfahren der Mathematiktests. Zunächst werden in der folgenden Tabelle die zugrundeliegenden Fallzahlen berichtet, gefolgt von den Angaben zu den minimal erzielten Rohwertpunkten pro Subskala und den maximal erzielten Rohwertpunkten. Darüber hinaus werden Mittelwerte, Standardabweichung und die Angaben zu Schiefe und Kurtosis dargestellt.

Tabelle 9

Deutscher Mathematiktest (DEMAT) für die vier Messzeitpunkte

	<i>N</i>	<i>Min</i>	<i>Max</i>	<i>M</i>	<i>SD</i>	σ	<i>Schiefe</i>	<i>Kurtosis</i>
Messzeitpunkt 1								
DEMAT 1 gesamt	1093	2	36	25.82	8.43	.25	-.98	.15
Messzeitpunkt 2								
DEMAT 2 gesamt	1231	0	36	20.47	9.85	.28	-.41	-.98
Messzeitpunkt 3								
DEMAT 3 gesamt	1277	0	31	17.94	6.02	.17	-.40	-.22
Messzeitpunkt 4								
DEMAT 4 gesamt	1067	0	39	21.60	7.32	.22	.02	-.68

N = Stichprobengröße; *Min* = Minimum; *Max* = Maximum; *M* = Mittelwert; *SD* = Standardabweichung; σ = Standardfehler.

Die Normalverteilungsprüfung für das DEMAT 1-Testverfahren zeigt eine leichte Abweichung von der Normalverteilung (Schiefe -.98), die jedoch insgesamt zu rechtfertigen ist. Für den DEMAT 2 ist eine leichte Steilheit festzustellen (Kurtosis -.98), während die Verfahren DEMAT 3 und DEMAT 4 normalverteilt sind. Nachfolgende Abbildungen geben

die Verteilung der Rohwerte für die Gesamttestverfahren grafisch aus, während die Normalverteilungsprüfung für die insgesamt 37 Subskalen der Mathematikleistungstests im Anhang auf der CD hinterlegt sind.

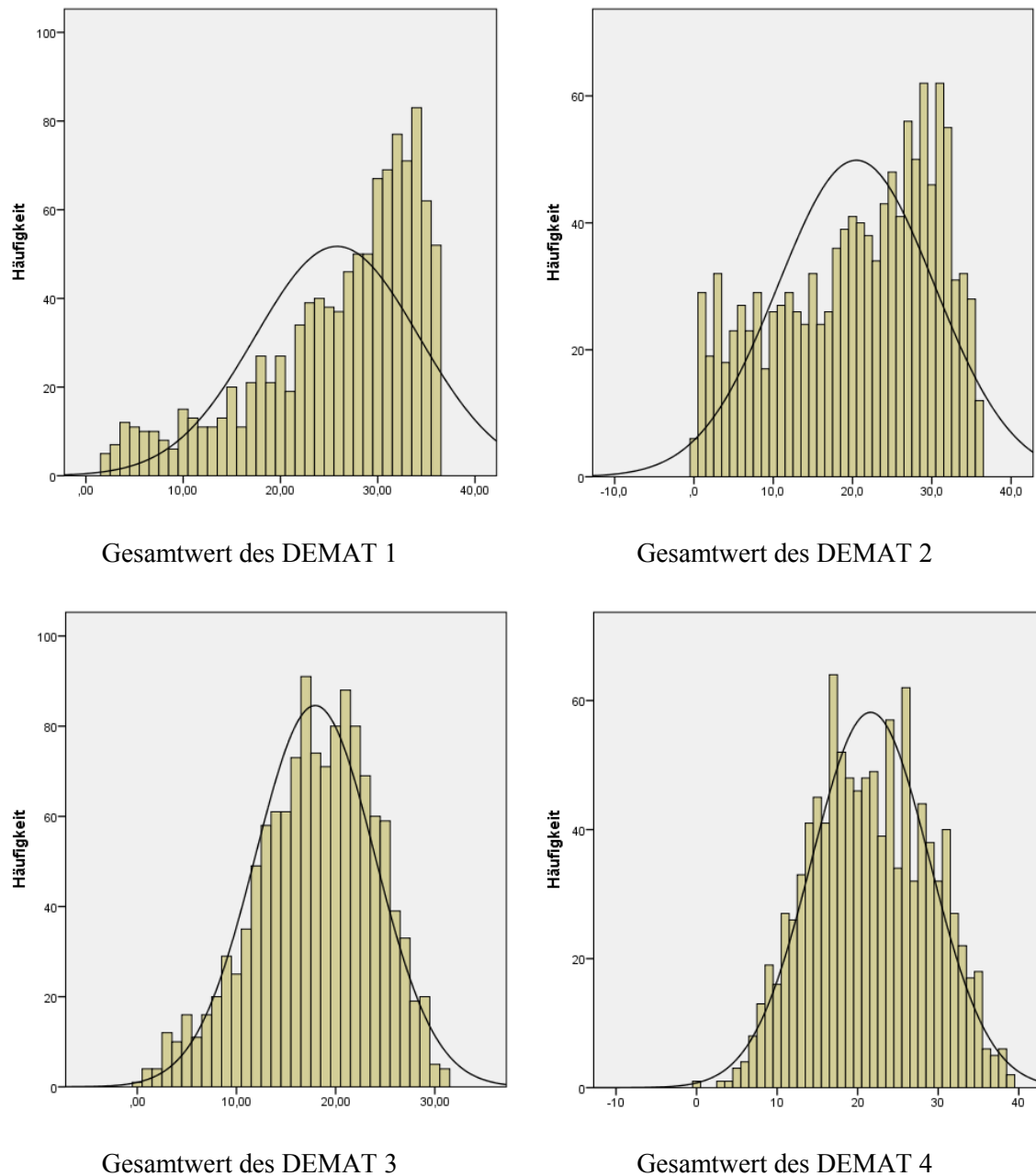


Abbildung 30 Verteilung der Rohwerte im DEMAT (Gesamtwerte)

Insgesamt erlauben die Verteilungsparameter für die eingesetzten Verfahren die Analysen. Im folgenden Abschnitt sollen zunächst die Korrelationen auf manifester Ebene berichtet werden.

Korrelationen KFT und DEMAT auf manifester Ebene über vier Messzeitpunkte

Die folgenden beiden Tabellen zeigen die Korrelationen auf manifester Ebene über vier Messzeitpunkte.

Tabelle 10

Korrelationen über die vier Messzeitpunkte – manifeste Variablen I, analysierte Fälle $N = 1665$

KFT	1 V	1 Q	1 N	2 V	2 Q	2 N	3 V	3 Q	3 N	4 V	4 Q	4 N
1 V	1											
1 Q	0.66	1										
1 N	0.40	0.58	1									
2 V	0.66	0.57	0.40	1								
2 Q	0.57	0.66	0.50	0.71	1							
2 N	0.43	0.52	0.55	0.53	0.69	1						
3 V	0.50	0.55	0.43	0.56	0.61	0.53	1					
3 Q	0.40	0.54	0.46	0.44	0.60	0.56	0.66	1				
3 N	0.42	0.49	0.49	0.45	0.56	0.61	0.57	0.66	1			
4 V	0.49	0.58	0.48	0.57	0.61	0.55	0.75	0.62	0.55	1		
4 Q	0.37	0.51	0.46	0.42	0.57	0.52	0.56	0.66	0.55	0.68	1	
4 N	0.37	0.48	0.47	0.44	0.54	0.56	0.50	0.58	0.64	0.62	0.68	1

V = verbal; Q = quantitativ; N = nonverbal.

Für den verbalen Bereich des KFT 1 kann eine mittlere bis hohe Korrelation zu den anderen verbalen Skalen festgestellt werden, während auch substantielle Beziehungen zur quantitativen Subskala festzustellen sind. Bei der genaueren Betrachtung der Aufgaben dieser Skala fällt auch das erforderliche Sprachverständnis auf und erklärt somit die hohen Korrelationen (siehe Beschreibung der Testverfahren).

Tabelle 11

Korrelationen über die vier Messzeitpunkte – manifeste Variablen II, analysierte Fälle $N = 1665$

		DEMAT											
		1 G	1 A	1 S	2 G	2 A	2 S	3 G	3 A	3 S	4 G	4 A	4 S
KFT	1 V	.21	.35	.35	.25	.35	.34	.27	.31	.27	.31	.32	.35
	1 Q	.26	.49	.44	.33	.53	.46	.40	.50	.38	.42	.46	.50
	1 N	.22	.41	.36	.28	.44	.34	.38	.43	.32	.38	.40	.38
	2 V	.27	.44	.38	.28	.46	.40	.36	.39	.37	.39	.39	.42
	2 Q	.37	.56	.48	.37	.61	.54	.46	.52	.48	.50	.52	.56
	2 N	.29	.47	.43	.30	.44	.34	.40	.48	.40	.43	.46	.43
	3 V	.23	.49	.44	.26	.55	.47	.41	.54	.44	.43	.47	.55
	3 Q	.25	.58	.45	.31	.64	.52	.47	.62	.50	.45	.53	.58
	3 N	.20	.42	.34	.28	.42	.34	.43	.47	.38	.41	.43	.43
	4 V	.28	.57	.48	.31	.58	.51	.45	.56	.49	.44	.54	.61
	4 Q	.30	.60	.48	.34	.62	.51	.47	.62	.49	.47	.59	.63
	4 N	.27	.46	.38	.36	.52	.40	.46	.51	.44	.49	.47	.50

V = verbal; Q = quantitativ; N = nonverbal; G = Geometrie; A = Arithmetik; S = Sachrechnen.

Die Korrelationen der Subskala KFT 1 „verbal“ zu allen DEMAT Skalen sind deutlich geringer. Die Korrelationen der quantitativen Skala des KFT 1 ist über alle Messzeitpunkte in mittlerer bis deutlicher Höhe und es bestehen auch Korrelationen zum nonverbalen Bereich des KFT 2, was durch die konzeptionelle Orientierung dieses Bereichs an einer eher allgemeinen kognitiven Fähigkeit begründet wird (Abou-Koura & Perleth, 2005; Heller & Perleth, 2000; Perleth & Heller, 2008). Auch auf der manifesten Ebene sind die Korrelationen

dieser Skala zu Skalen des KFT höher als zu den Skalen des Mathematikleistungstestverfahren zu den Messzeitpunkten, was auch auf manifester Ebene nicht für eine Identität der Konstrukte spricht. Höhere Korrelationen (.62) sind zwischen den Skalen des KFT „quantitativ“ und der Skala Arithmetik des DEMAT 3 festzustellen. Auch die quantitative Skala des KFT 4 korreliert mit .63 deutlich mit der quantitativen Skala des DEMAT 4. Letztgenannte Skala zeigt wiederum auch deutliche Korrelationen zum nonverbalen Bereich des KFT 4, was auf manifester Ebene für eine Überlappung der erforderlichen Fähigkeiten zur Lösung beider Aufgabenbereiche spricht. Auch die engen Beziehungsn des KFT 4 „quantitativ“ zur nonverbalen Subskala des KFT, welche vornehmlich allgemeine kognitive Fähigkeiten fokussiert, unterstreichen dies.

Die Korrelationen im Längsschnitt über die vier Messzeitpunkte im jährlichen Abstand auf manifester Ebene zeigen Korrelationen in mittlerer Höhe. Für die quantitativen Skalen des KFT 1 und 2 zeigen sich recht hohe Korrelationen in einer Höhe von .66 bzw. .71 zu den verbalen Skalen des KFT 1 und 2. Auch der Zusammenhang zwischen der nonverbalen Skala des KFT 3 und der quantitativen Skala des KFT 2 ist recht hoch mit einem Wert von .69. Auch für die Verfahren für die dritte Klassenstufe zeigen sich ausgeprägte Zusammenhänge zwischen den quantitativen Subskalen und den verbalen Subskalen in einer Höhe von .66 und im KFT 4 in Höhe von .68.

Korrelationen manifester Variablen der DEMAT (Subskalen)

In der untenstehenden Tabelle sind die manifesten Korrelationen der Subskalen Arithmetik (AR), Sachrechnen (SR) und Geometrie (GO) untereinander über die Messzeitpunkte angegeben.

Tabelle 12

Korrelationen zwischen den manifesten Variablen (standardisiert), DEMAT – Subskala „Geometrie“

	DEMAT 1	DEMAT 2	DEMAT 3	DEMAT 4
DEMAT 1	1.00			
DEMAT 2	.18	1.00		
DEMAT 3	.12	.33	1.00	
DEMAT 4	.17	.33	.50	1.00

Die Korrelation der Subskalen des 3. und 4. Messzeitpunktes ist hierbei am höchsten, gefolgt von den Zusammenhängen des 2. und 4. bzw. nachfolgend des 2. und 3. DEMAT-Testverfahrens. Am geringsten sind die Korrelationen zwischen der Subskala Geometrie des Testverfahrens für die 1. Klassenstufe zu den Testverfahren der 2. und nachfolgend der 3. Klassenstufe.

Für die Korrelationen des Subtests Arithmetik (siehe folgende Tabelle) ergibt sich folgendes Bild. Insgesamt zeigt sich, dass die Skalenzusammenhänge in der Arithmetik über die Testverfahren der 1. bis 4. Klassenstufe hinweg sehr hoch sind. So besteht zwischen den Arithmetikskalen des DEMAT 2 und des DEMAT 3 der höchste Zusammenhang. Auch die Arithmetikskala des DEMAT 1 und DEMAT 2 weisen deutliche Zusammenhänge aus. Insgesamt zeigen sich über alle Verfahren hinweg sehr zufriedenstellende Korrelationen.

Tabelle 13

Korrelationen zwischen den manifesten Variablen (standardisiert), DEMAT – Subskala „Arithmetik“

	DEMAT 1	DEMAT 2	DEMAT 3	DEMAT 4
DEMAT 1	1.00			
DEMAT 2	.62	1.00		
DEMAT 3	.54	.62	1.00	
DEMAT 4	.51	.56	.59	1.00

Die Korrelationen der Subskalen Sachrechnen (Tabelle unten) zeigen ebenso recht zufriedenstellende Zusammenhänge über die DEMAT 1-4 Verfahren.

Tabelle 14

Korrelationen zwischen den manifesten Variablen (standardisiert), DEMAT – Subskala „Sachrechnen“

	DEMAT 1	DEMAT 2	DEMAT 3	DEMAT 4
DEMAT 1	1.00			
DEMAT 2	.42	1.00		
DEMAT 3	.30	.38	1.00	
DEMAT 4	.43	.49	.43	1.00

Korrelationen der KFT-Subskalen auf manifester Ebene

Die folgende Tabelle zeigt die Korrelationen der Subskalen des KFT in den Bereichen quantitative, verbale und nonverbale Fähigkeiten untereinander über die Messzeitpunkte.

Tabelle 15

Korrelationen zwischen den manifesten Variablen (standardisiert), KFT – Subskala „quantitativ“

	KFT 1	KFT 2	KFT 3	KFT 4
KFT 1	1.00			
KFT 2	.66	1.00		
KFT 3	.53	.60	1.00	
KFT 4	.50	.56	.65	1.00

Die Korrelationen der Subskalen des 1. und 2. Messzeitpunktes sind hierbei am höchsten, gefolgt von den Zusammenhängen des 3. und 4. bzw. nachfolgend des 2. und 3. KFT-Test-

verfahrens. Am geringsten sind die Korrelationen zwischen der Subskala quantitative Fähigkeiten des Testverfahrens für die 1. Klassenstufe und dem Testverfahren der 3. Klassenstufe, wobei diese dennoch im mittleren Bereich angesiedelt sind (siehe folgende Tabelle).

Für die Korrelationen des Subtests verbale Fähigkeiten (Tabelle unten) ergibt sich folgendes Bild. Insgesamt zeigt sich, dass die Skalenzusammenhänge über die Testverfahren der 1. bis 4. Klassenstufe hinweg hoch sind. So besteht zwischen den KFT 4-Skalen und den Skalen des KFT 3 der höchste Zusammenhang (Stabilitätshinweis gegen Ende der Grundschulzeit). Aber auch zwischen den Verbalskalen des KFT 1 und KFT 2 bestehen recht hohe Zusammenhänge. Die Zusammenhänge der verbleibenden Skalen sind ausnahmslos im mittleren Bereich. Insgesamt zeigen sich über alle Skalen der verbalen Tests über die Jahre hinweg sehr zufriedenstellende Korrelationen.

Tabelle 16

Korrelationen zwischen den manifesten Variablen (standardisiert), KFT – Subskala „verbal“

	KFT 1	KFT 2	KFT 3	KFT 4
KFT 1	1.00			
KFT 2	.66	1.00		
KFT 3	.49	.55	1.00	
KFT 4	.49	.56	.74	1.00

Die Korrelationen der Subskalen nonverbal, aufgeführt in der folgenden Tabelle, zeigen ebenso sehr zufriedenstellende Zusammenhänge über die KFT 1-4-Verfahren.

Tabelle 17

Korrelationen zwischen den manifesten Variablen (standardisiert), KFT – Subskala „nonverbal“

	KFT 1	KFT 2	KFT 3	KFT 4
KFT 1	1.00			
KFT 2	.54	1.00		
KFT 3	.44	.61	1.00	
KFT 4	.46	.55	.63	1.00

So zeigen sich wiederum die stärksten Zusammenhänge zwischen dem KFT 3 und dem KFT 4, jedoch sind durchweg mittelstarke Zusammenhänge zwischen den nonverbalen Skalen zwischen dem KFT 1 und dem KFT 2 und fast mittlere Zusammenhänge zwischen dem KFT 3 und dem KFT 4 zu verzeichnen. Die Stabilität von KFT 2 nach KFT 3 kann mit einer mittleren Korrelation von .61 belegt werden, sowie einer mittleren Korrelation zwischen dem KFT 2 und dem KFT 4 von .55.

4.1 Ergebnisse 1 – Faktorielle Differenziertheit

Für die Beantwortung der Fragestellung 1 zur Konstruktunabhängigkeit wurde ein Latent-State-Modell mit Berücksichtigung der Korrelationen der nicht beobachtbaren Messfehler und situationsspezifischer Einflüsse mit latenten Variablen, Faktorladungen, Residualvarianzen und Residualkorrelationen spezifiziert. Zunächst wurden die Korrelationen auf der manifesten Ebene, dann auf latenter Ebene geprüft.

4.1.1 Korrelationen manifeste Faktoren auf Subtestebene:

KFT 1, KFT 2 – DEMAT 1, DEMAT 2

In diesem Abschnitt wird nun der faktoriellen Differenziertheitsfrage auf manifester Ebene zunächst für den ersten Messzeitpunkt nachgegangen. Die Konstrukte, die den Messverfahren KFT und DEMAT für die erste Klassenstufe zugrunde liegen, werden zuerst verglichen, während dann die Betrachtung der Konstruktunabhängigkeit auf manifester Ebene für den zweiten Messzeitpunkt erfolgt.

Die Tabelle zeigt die Korrelationen der Konstrukte untereinander, sowie die Korrelationsbeziehungen zum jeweils anderen Konstrukt zum selben Messzeitpunkt. Es wird deutlich, dass dieselben Konstrukte untereinander jeweils höher korrelieren. Zum jeweils anderen Konstrukt bestehen die höchsten Korrelationen zum ersten Messzeitpunkt zwischen der Subskala Arithmetik und dem quantitativen Teil des KFT mit .49.

Zum zweiten Messzeitpunkt besteht der höchste Zusammenhang zwischen dem quantitativen Teil des KFT und der Arithmetiksubskala mit .61. Über die Zeit von zwei Jahren gesehen sind die Korrelationen im mittleren bis höheren Bereich. So ist die Beziehung zwischen Arithmetik des ersten Messzeitpunktes und dem Sachrechnen des zweiten Messzeitpunktes recht hoch ausgeprägt. Es wird deutlich, dass dieselben Konstrukte untereinander jeweils höher korrelieren, was auf die größeren Zusammenhänge desselben Konstrukts hinweist. Zum anderen Konstrukt bestehen die höchsten Korrelationen zum ersten Messzeitpunkt zwischen den Mathematikleistungen des Bereichs der Arithmetik und dem quantitativen Teil des kognitiven Fähigkeitstests mit .49.

Tabelle 18

Korrelationen zum ersten und zweiten Messzeitpunkt – manifeste Faktoren

Modell	KFT 2 verbal	KFT 2 quantitativ	KFT 2 nonverbal	DEMAT 2 Geometrie	DEMAT 2 Arithmetik	DEMAT 2 Sachrechnen
KFT 1 verbal	1	.71	.53	.28	.46	.39
KFT 1 quantitativ	.66	1	.69	.37	.61	.54
KFT 1 nonverbal	.40	.58	1	.30	.44	.35
DEMAT 1 Geometrie	.21	.26	.22	1	.39	.32
DEMAT 1 Arithmetik	.35	.49	.41	.36	1	.67
DEMAT 1 Sachrechnen	.35	.44	.36	.29	.62	1

Zum zweiten Messzeitpunkt besteht der höchste Zusammenhang zwischen dem quantitativen Teil des kognitiven Fähigkeitstests und dem Arithmetiksubtest des DEMAT- Testverfahrens mit .61.

4.1.2 Korrelationen manifester Faktoren auf Subtestebene

KFT 3, KFT 4 – DEMAT 3, DEMAT 4

Diese folgende Tabelle zeigt die Korrelationen der manifesten Variablen zur Beantwortung der Unabhängigkeitsfrage der Konstrukte kognitive Fähigkeiten und mathematische Schulleistungen für das Verfahren für den dritten Messzeitpunkt und für das Verfahren für die vierte Klassenstufe (vierten Messzeitpunkt).

Tabelle 19

Korrelationen zum dritten und vierten Messzeitpunkt – manifeste Faktoren

Modell	KFT 4 verbal	KFT 4 quantitativ	KFT 4 nonverbal	DEMAT 4 Geometrie	DEMAT 4 Arithmetik	DEMAT 4 Sachrechnen
KFT 3 verbal	1	.68	.62	.44	.54	.61
KFT 3 quantitativ	.66	1	.68	.47	.59	.63
KFT 3 nonverbal	.56	.66	1	.49	.47	.50
DEMAT 3 Geometrie	.41	.47	.38	1	.40	.47
DEMAT 3 Arithmetik	.54	.62	.47	.48	1	.57
DEMAT 3 Sachrechnen	.44	.50	.43	.47	.48	1

Für die manifesten Variablen des dritten Messzeitpunktes sind die Korrelationen zum jeweiligen Subtest des gleichen Konstrukts höher als zum anderen Konstrukt. Zugleich aber

lassen sich auch jeweils die Zusammenhänge zwischen den Konstrukten kognitive Fähigkeiten und mathematische Schulleistungen erkennen.

4.1.3 Faktorielle Differenziertheit (latent) – KFT 1 vs. DEMAT 1

Nachfolgend werden die Ergebnisse der faktoriellen Differenziertheit der Konstrukte berichtet. Zunächst erfolgt wie in der folgenden Abbildung der Vergleich der Konstrukte zum ersten Messzeitunkt.

Die Faktorladungen des KFT-Gesamtwertes der latenten Variablen laden mit .94 am höchsten und mit einem Wert von .62 auf die nonverbale Skala am niedrigsten. Insgesamt ist der Wert aber dennoch recht hoch. Die Faktorladung auf den verbalen Bereich ist mit .70 wiederum höher. In der Betrachtung der latenten Variablen des DEMAT-Gesamtwertes auf die Arithmetiksubskala ist mit .84 am höchsten und mit .42 auf die Skala Sachrechnen ebenfalls recht hoch. Am niedrigsten ist die Faktorladung auf die Geometriesubskala mit .42. Die Unabhängigkeitsprüfung der untersuchten Konstrukte kognitive Fähigkeiten, gemessen mit dem kognitiven Testverfahren KFT für die erste Klassenstufe und die mathematischen Schulleistungen, die mit dem DEMAT-Testverfahren für die erste Klassenstufe erhoben worden sind, zeigen eine Korrelation auf der latenten Ebene von .65 (siehe folgende Abbildung). Auf der latenten Ebene sind die Korrelationen zwischen den Konstrukten kognitive Fähigkeiten und mathematische Schulleistungen in mittlerer Höhe, was darauf hinweist, dass die Konstrukt Faktoren unterschieden werden können.

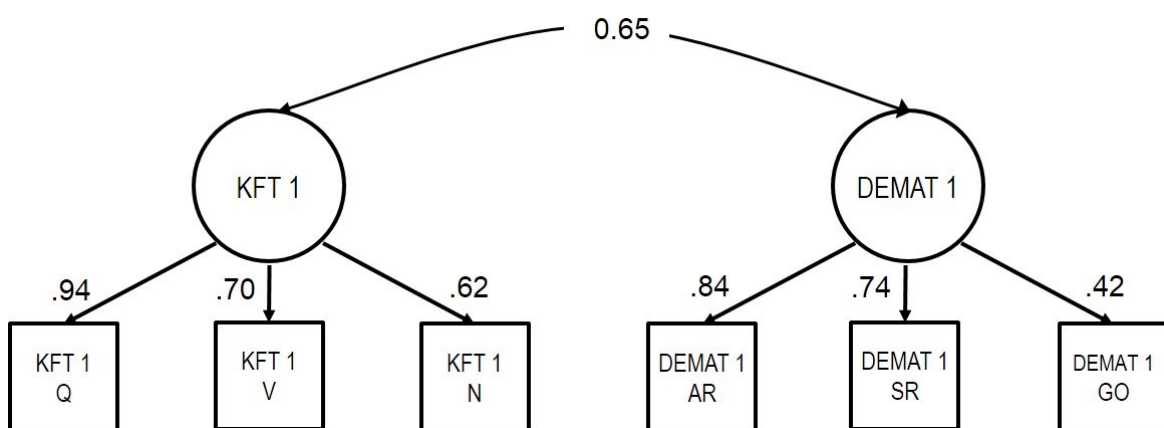


Abbildung 31 Faktorielle Differenziertheit KFT 1 vs. DEMAT 1;
V = verbal; Q = quantitativ; N = nonverbal.

Im Folgenden sollen die Konstrukte, die die kognitiven Fähigkeiten und die mathematischen Schulleistungen für die zweite Klassenstufe messen, dargestellt werden.

4.1.4 Faktorielle Differenziertheit (latent) – KFT 2 vs. DEMAT 2

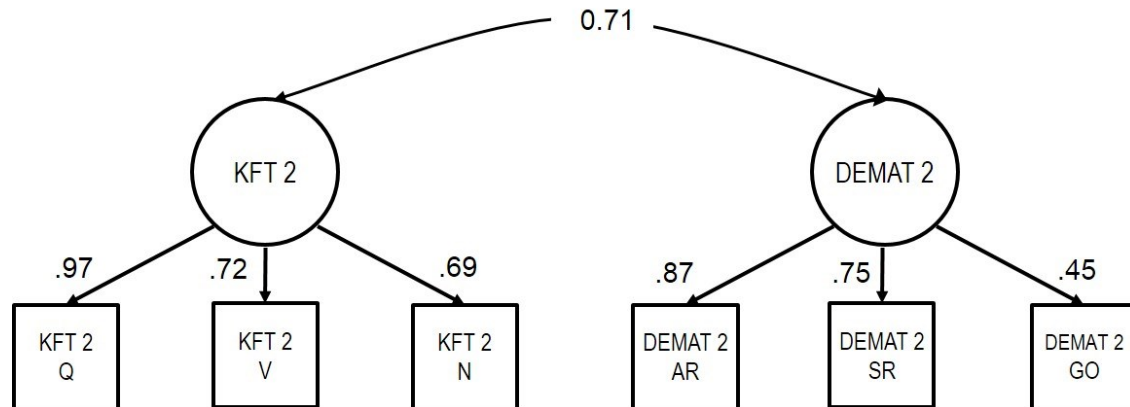


Abbildung 32 Faktorielle Differenziertheit KFT 2 vs. DEMAT 2;
AR = Arithmetik; GO = Geometrie; SR = Sachrechnen; V = verbal; Q = quantitativ;
N = nonverbal.

Die Prüfung der faktoriellen Differenziertheit für die Konstrukte, die mit dem KFT 2 und swm DEMAT 2 erfasst werden, ergab folgende Ergebnisse, die in Abbildung 44 veranschaulicht werden.

Die Faktorladungen der latenten KFT-Variablen laden am deutlichsten mit einem Wert von .92 auf den quantitativen Bereich des KFT und am geringsten auf die nonverbale Skala des KFT mit einem Wert von .69, während die Faktorladung auf die verbale Skala mit .72 lädt. Die Faktorladungen der latenten DEMAT-Variablen auf die jeweiligen Subskalen sind mit .87 auf die Arithmetikskala am höchsten und mit .45 auf die Geometrieskala am niedrigsten, während wiederum die Faktorladung auf die Skala Sachrechnen mit .75 einen deutlich höheren Wert aufweist.

In dem dargestellten Modell sind die latenten Korrelationsbeziehungen zwischen dem Konstrukt kognitive Fähigkeiten, das mit dem KFT gemessen wird, und dem Konstrukt, welches mit dem Deutschen Mathematikleistungstest der zweiten Klassenstufe gemessen wird, abgebildet. Die dargestellten latenten Variablen des jeweils anderen Konstrukts korrelieren mit .71 im mittleren Bereich miteinander, woraus geschlossen werden kann, dass auf der nicht beobachtbaren Ebene die gleichen Einflüsse wirken. Die Messfehler

korrelieren innerhalb der latenten Variablen und deren Mittelwerten. Sie stellen die Messfehler dar, die über die Messzeitpunkte wiederholt auftreten und dürfen somit miteinander korrelieren.

4.1.5 Faktorielle Differenziertheit (latent) – KFT 3 vs. DEMAT 3

Für die Prüfung der faktoriellen Differenziertheit der Messung der Konstrukte, die mit dem KFT 3 erfasst werden und dem Konstrukt mathematische Schulleistungen, das mit dem DMEAT 3 ermittelt wird, zeigte folgendes Ergebnis. Die unten dargestellte Grafik zeigt die Ergebnisse der Fragestellung 4 für die Unabhängigkeit der Konstrukte in der 3. Klassenstufe. Die Korrelationen auf der latenten Ebene weisen hier eine recht hohe Korrelationen mit .87 zwischen den latenten Variablen der jeweiligen anderen Konstrukte aus und zeigen, dass die latenten Beziehungen damit konstant über die Zeit bestehen.

Die Faktorladungen der latenten Variablen des KFT 3 laden mit .88 am höchsten auf den quantitativen Bereich, mit .75 auf die verbalen Subskalen und mit .73 am geringsten, aber insgesamt dennoch hoch, auf die nonverbale Subskala. Die Faktorladungen der latenten Variablen des DEMAT zeigen insgesamt etwas niedrigere Werte und laden innerhalb des Konstrukts mit .75 am höchsten auf den Arithmetikbereich, mit .64 deutlich niedriger auf die Skala Sachrechnen und mit annähernd gleicher Ausprägung auf den Geometriebereich.

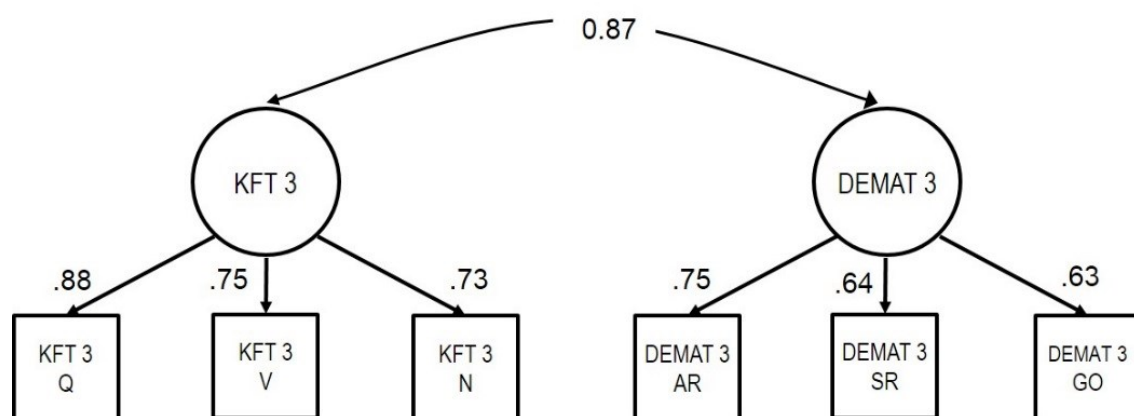


Abbildung 33 Faktorielle Differenziertheit KFT 3 vs. DEMAT 3;
 AR = Arithmetik; GO = Geometrie; SR = Sachrechnen; V = verbal; Q = quantitativ;
 N = nonverbal.

Im Folgenden soll abschließend die faktorielle Differenziertheit der Testverfahren für die vierten Klassenstufen ermittelt werden.

4.1.6 Faktorielle Differenziertheit (latent) – KFT 4 vs. DEMAT 4

Die dargestellte Grafik zeigt die Ergebnisse der Fragestellung 4 für die Unabhängigkeit der Konstrukte der Messung der Konstrukte für die vierte Klassenstufe.

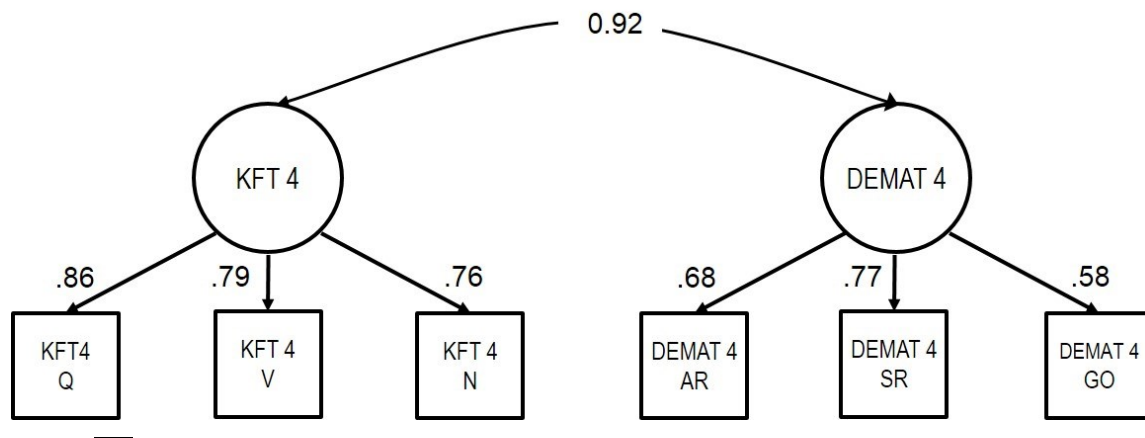


Abbildung 34 Faktorielle Differenziertheit KFT 4 vs. DEMAT 4;
AR = Arithmetik; GO = Geometrie; SR = Sachrechnen; V = verbal; Q = quantitativ;
N = nonverbal.

Die Faktorladungen des kognitiven Fähigkeitstests der vierten Klassenstufe auf den quantitativen Bereich des Konstrukts sind mit .86 am höchsten, während die Faktorladung des KFT-Gesamtwertes auf den verbalen Bereich mit .79 etwas geringer, aber weiterhin hoch ist. Die Faktorladung des Gesamtergebnisses des KFT auf die nonverbale Subskala ist mit .76 zwar am geringsten, aber in der Gesamtbetrachtung dennoch als hoch anzusehen. Für das Modell der mathematischen Schulleistungen für die vierten Klassenstufen (4. Messzeitpunkt) zeigen sich im Vergleich zu den Werten des KFT insgesamt niedrigere Faktorladungen, die am geringsten auf den Geometriebereich laden, während der Bereich des Sachrechnens mit .77 am höchsten lädt und die Faktorladung auf latenter Ebene auf den Arithmetikbereich mit .68.

Auf der latenten Ebene wird deutlich, dass auch die Korrelationen zwischen den latenten Variablen der gegenteiligen Konstrukte mit einem Wert von .92 hoch sind und damit konstant über die Zeit bestehen.

Kontrolle der Mehrebenenstruktur in Modellen zur Prüfung der faktoriellen Differenziertheit

Die Unabhängigkeitsprüfung wurde nochmals unter Kontrolle der Mehrebenenstruktur vorgenommen. Dabei wurde die Clusterung der Daten innerhalb der verschiedenen Schulen berücksichtigt. Die untenstehende Tabelle zeigt im Vergleich der Daten mit Berücksichtigung der Mehrebenenstruktur kaum Unterschiede zu den Analyseergebnissen der Modelle ohne Berücksichtigung der Mehrebenenstruktur.

Der χ^2 -Wert zum ersten Messzeitpunkt erhöht sich unter Berücksichtigung der Mehrebenenstruktur gering auf 32.30. In der Betrachtung des zweiten Messzeitpunktes ist er im Vergleich der Modelle ohne Berücksichtigung der Mehrebenenstruktur zum Modell mit Berücksichtigung der Mehrebenenstruktur mit 16.61 jedoch deutlich geringer. Zum dritten Messzeitpunkt ist der χ^2 -Wert unter Berücksichtigung der Schulclusterung mit einem Wert von 29.35 nur geringfügig geringer als im Modell ohne Berücksichtigung der Mehrebenenstruktur, das einen χ^2 -Wert von 30.36 aufweist. Für den vierten Messzeitpunkt ist der χ^2 -Wert mit 46.35 unter Berücksichtigung der Mehrebenenstruktur geringer als im Modell ohne Berücksichtigung der Mehrebenenstruktur, das einen χ^2 -Wert von 46.35 aufweist. Die Anzahl der Freiheitsgrade bleibt konstant mit 8, während der CFI mit Werten zwischen .98 und .99 für die Modelle ohne Berücksichtigung der Mehrebenenstruktur sehr zufriedenstellende Werte aufweist und damit von einer guten Passung auf die vorliegende Datenstruktur hinweist. Der CFI für die Modelle unter Kontrolle der Mehrebenenstruktur weisen dem niedrigsten Wert von .99 zum 1. Messzeitpunkt und dem höchsten Wert von .99 zum zweiten Messzeitpunkt eine ebenso gute Passung auf die Daten aus. Der RMSEA zeigt im Modellvergleich ohne bzw. mit Berücksichtigung der Mehrebenenstruktur zum 1. Messzeitpunkt so gut wie keine Änderung und liegt mit Werten von 0.99 (ohne Mehrebenenstruktur) und .99 (mit Berücksichtigung der Mehrebenenstruktur) im akzeptablen Bereich.

Tabelle 20

Passungswerte angewandter Modelle zur Prüfung der Unabhängigkeit der Konstrukte der 4 Messzeitpunkte jeweils ohne und mit Kontrolle der Mehrebenenstruktur

Modell	Messzeitpunkt							
	1 ohne MEK	1 mit MEK	2 ohne MEK	2 mit MEK	3 ohne MEK	3 mit MEK	4 ohne MEK	4 mit MEK
χ^2	32.07	32.30	26.71	16.61	30.36	29.35	51.78	46.35
<i>df</i>	8	8	8	8	8	8	8	8
<i>CFI</i>	.99	.99	.99	.99	.99	.99	.98	.99
<i>RMSEA</i>	.05	.04	.04	.03	.05	.06	.06	.06
<i>SRMR</i>	.035	.03	.02	.02	.02	.02	.02	.02
<i>LS kor</i>	.65	.65	.71	.70	.87	.87	.92	.92

MEK = Mehrebenenkontrolle; χ^2 = Diskrepanz; *df* = Freiheitsgrade; *CFI* = Comparative Fit Index; *RMSEA* = Root Mean Square Error of Approximation; *LS kor* = Korrelation der Latent-State-Variablen.

Im Modellvergleich zum zweiten Messzeitpunkt zeigt sich im Modell unter Kontrolle der Mehrebenenstruktur eine geringere Ausprägung mit .03 im Vergleich zum Modell ohne Berücksichtigung der Mehrebenenstruktur. Zum dritten Messzeitpunkt ist der RMSEA unter Kontrolle der mit Mehrebenenstruktur etwas höher als im Modell ohne Berücksichtigung der Mehrebenenstruktur .06. Der SRMR zeigt über die betrachteten Zeitpunkte im Modellvergleich derselben Messzeitpunkte mit Berücksichtigung der Mehrebenenstruktur und dem jeweiligen Modell ohne Berücksichtigung der Mehrebenenstruktur keine Änderung. Die Sichtung der Analyseergebnisse für die Latent-State-Korrelationen zeigt ebenso keine Änderung der Werte, wenn entweder die Mehrebenenstruktur kontrolliert bzw. nicht kontrolliert wurde.

4.1.7 Faktorielle Differenziertheit – Zusammenfassung und Diskussion der Ergebnisse

Mittels der Spezifikation über Latent-State-Modelle ist es gelungen, die Konstrukte kognitive Fähigkeiten und mathematische Schulleistungen auf Subtestebene präzise empirisch über die Zeit zu überprüfen.

Die Ergebnisse sind besonders aussagekräftig, da für jeden Subtest eine große Anzahl an Items zur Verfügung stand: Für den KFT des 1. Messzeitpunktes standen 72 Items zur Verfügung, für den 2. Messzeitpunkt des KFT ebenfalls 72 Items, und für den KFT 3 140 Items, sowie für den KFT 4 195 Items. Für die Erfassung mathematischer Schulleistungen standen zum 1. Messzeitpunkt 36 Items zur Verfügung und für den 2. Messzeitpunkt 57. Der DEMAT 3 erfasste 37 Items und zum 4. Messzeitpunkt wurden 40 Items erfasst.

Es wurde durch die Spezifikation von Latent-State-Modellen eine klare Trennung der Konstrukte kognitive Fähigkeiten und Mathematikleistungen über die vier Messzeitpunkte nachgewiesen.

Die Korrelationen der latenten State-Variablen kognitiver Fähigkeiten und mathematischer Schulleistungen zeigen Zusammenhänge über alle vier Messzeitpunkte im mittleren Bereich hinweg. Die Faktorladungen des jeweils selben Konstrukts zugleich laden höher. Daher kann geschlussfolgert werden, dass die gleichen Einflüsse wirksam und als Stabilitätsindikatoren hinsichtlich interindividueller Unterschiede eines Konstrukts interpretierbar sind.

Es kann also geschlussfolgert werden, dass die faktorielle Differenziertheit kognitiver Fähigkeiten (gemessen mit dem KFT-Testverfahren) und mathematischer Schulleistungen (gemessen mit dem DEMAT-Testverfahren) sowohl auf manifester als auch auf latenter Ebene empirisch belegbar ist.

Die auch auf latenter Ebene nachgewiesene faktorielle Differenziertheit der erfassten Konstrukte (kognitive Fähigkeiten und mathematische Leistungen) kann jene theoretischen Argumentationen empirisch stützen, die nicht von identischen, sondern von präzise trennbaren, unabhängigen Bereichen ausgehen. Sie stimmen diesbezüglich mit den Ergebnissen von Soares (2015) überein, dass es sich bei beiden Konstrukten um unabhängige Konstrukte handelt.

Mit der Messung kognitiver Fähigkeiten und der mathematischen Schulleistungen werden unabhängige Konstrukte erfasst. Mit diesem empirischen Nachweis der präzisen

Trennbarkeit der Konstrukte können diese Ergebnisse einen theoretischen Beitrag zu den bislang heftig diskutierten Theoriekontroversen leisten.

Es lässt sich aus ihnen schließen, dass mathematische Schulleistungen das Ergebnis vorheriger schulischer Lernprozesse sind. Daraus ergeben sich Konsequenzen für die Lernforschung. Diese Ergebnisse stehen im Einklang mit den Forschungsergebnissen von Geary (1995) und Köller und Baumert (2002). Doch auch jene theoretischen Befunde können gestützt werden, die von der Trainierbarkeit kognitiver Fähigkeiten ausgehen, wie beispielsweise von Floyd et al. (2013) dokumentiert, denn mathematische Schulleistungen besitzen einen größeren Einfluss auf kognitive Fähigkeiten als umgekehrt. Darüber hinaus ist der Einfluss mathematischer Schulleistungen im Laufe der Grundschulzeit zunehmend.

Die gegenseitige Vermittlungsrolle mathematischer Schulleistungen und kognitiver Fähigkeiten kann durch die Investitionstheorie von Cattell (1971) gestützt werden. Dementsprechend können mathematische Schulleistungen förderliche Wirkung auf kognitive Fähigkeiten aufweisen, während kognitive Fähigkeiten gleichermaßen positiv auf mathematische Schulleistungen wirken. Die Rolle der kognitiven Fähigkeiten auf die schulische Leistungserbringung und die unterstützende Wirkung bei dem weiteren Wissenserwerb kann in Übereinstimmung mit Primi et al. (2010) bestätigt werden. Das früher erworbene Wissen (Vorwissen), das in dieser Arbeit als mathematische Schulleistungen erhoben wurde, zeigt im Grundschulalter über die betrachteten vier Messzeitpunkte einen zunehmenden Einfluss auf die kognitiven Leistungen. Diese werden bereichsspezifisch unterschiedlich beeinflusst und können als kristallisierte Intelligenzanteile verstanden werden. Diese wiederum, so zeigen die Ergebnisse der vorliegenden Studie, gehen als kognitives Potential wiederum in den weiteren Wissenserwerb ein. Damit kann die gegenseitige Beziehung beider Konstrukte bestätigt werden. Die Ergebnisse stützen damit auch die bisher vorliegenden Befunde, die den Einfluss des Vorwissens betonen (Perleth, 1997; Weinert & Helmke, 1997; Watkins et al., 2007; Krajewski & Schneider, 2006; Krajewski, 2008). Die vorliegenden Ergebnisse dieser Arbeit bestätigen den positiven Einfluss schulischer Bildung in Übereinstimmung mit den theoretischen Befunden, liefern jedoch präzisere Angaben über den zeitverzögerten Verlauf der Beziehungen im Grundschulalter. Da jedoch nachgewiesen werden konnte, dass standardisiert erfasste, stoffgebietsspezifisch erhobene mathematische Schulleistungen auch auf nonverbale Bereiche kognitiver Fähigkeiten wirken, welche konzeptionell als allgemeiner g-Faktor assoziiert verstanden werden, kann zudem davon ausgegangen werden, dass auch fluide kognitive Fähigkeiten von schulisch erworbenem Wissen beeinflusst werden.

Hinsichtlich der nicht unwesentlichen Debatten in der psychologischen Forschung, inwiefern es sich bei der Messung von kognitiven Fähigkeiten und Mathematikleistungen um separate, identische oder überlappende Konstrukte handelt, konnte mit dieser Arbeit insofern ein Beitrag zur Aufklärung geleistet werden, als dass sich präzise trennbare Faktoren extrahieren ließen, die die spezifische Verortung kognitiver Fähigkeiten zum einen und mathematischer Fähigkeiten zum anderen unterstützen. Somit kann die theoretische Ansicht einer eindeutigen Abgrenzbarkeit der Konstrukte unterstützt werden. Darüber hinaus zeigten sich auch innerhalb des Konstrukts mathematische Fähigkeiten bereichsspezifische Denk- und Wissensleistungen, die in unterschiedlicher Gewichtung für das jeweilige Leistungsniveau erforderlich sind. Für das Konstrukt kognitive Leistungen kann ebenso die theoretische Ansicht der Betrachtung der Bereichsspezifität kognitiver Faktoren unterstützt werden, deren Nachweis durch spezifische Faktoren belegt werden konnte.

Damit kann geschlussfolgert werden, dass eine differenzierte Analyse spezifischer kognitiver und auch wissensbasierter Leistungen spezifische bereichsspezifische Ressourcen sichtbar machen kann. Inwiefern sich daraus vielversprechende gezielte Interventionen ableiten lassen, die ansetzend an den spezifischen Bedürfnissen zur Potentialentfaltung ihre Wirkung auch auf andere (theoretisch trennbare) Bereiche entfalten können, kann Gegenstand weiterführender Forschung sein. Sofern dabei Leistungen im zeitlichen Verlauf im Vordergrund stehen, hängt die Aussagekraft von Resultaten erheblich von der Konstruktskonsistenz der in dem betrachteten Zeitraum eingesetzten Messinstrumente ab. Die vorliegende Arbeit kann daher auch zu weiterer Forschung in diesem Bereich anregen.

4.1.8 Zusammenfassung der Ergebnisse vor dem Hintergrund der Forschungsfragen

In der vorliegenden Arbeit stand die für die Forschungs- und pädagogische Praxis relevante Entwicklung kognitiver Fähigkeiten und mathematischer Leistungen im Grundschulalter im Mittelpunkt. Das Hauptziel war eine möglichst präzise Darstellung der komplexen Zusammenhänge zwischen kognitiven Fähigkeiten und Mathematikleistungen in der längsschnittlichen Betrachtung über vier Messzeitpunkte im jährlichen Abstand. Es wurden insgesamt vier Fragestellungen entwickelt. Die erste Fragestellung sucht eine grundlegende theoretische Antwort auf bislang kontroverse theoretische Debatten, während die drei folgenden Fragestellungen eher empirisch orientiert sind.

Da es eine erhebliche theoretische Diskussion darüber gibt, ob nicht beide Konstrukte kognitive Fähigkeiten auf der einen Seite und mathematische Leistungen auf der

anderen Seite, identische Fähigkeiten erfassen, sollte in der ersten Fragestellung der faktoriellen Differenziertheit der Konstrukte nachgegangen und geklärt werden, inwiefern es sich bei den Konstrukten kognitive Fähigkeiten und mathematische Leistung um präzise trennbare, unabhängige Bereiche handelt.

Die zweite empirische Fragestellung resultierte aus der Problematik schwieriger Vergleichbarkeitsbedingungen von Testleistungen von Grundschulkindern im Längsschnitt, da notwendige Testadaptionen an die Alterskohorte die Gefahr von kumulativen Messfehlern bergen. Daher sollte beantwortet werden, inwieweit die Konstrukte kognitive Fähigkeiten und mathematische Schulleistungen über den betrachteten Vierjahreszeitraum überhaupt vergleichbar sind.

Im Fokus der dritten Fragestellung stand, inwieweit die erfassten Konstrukte kognitive Fähigkeiten und mathematische Leistungen im Grundschulalter als stabil zu betrachten sind. Zur Stabilität kognitiver Leistungen und mathematischer Schulleistungen von Grundschulkindern im Längsschnitt existieren nur wenige Studien, die zudem nur zwei Messzeitpunkte berücksichtigen. Daher wurde diese noch offene Frage in den Fokus der vorliegenden Arbeit gerückt, die vier Messzeitpunkte im Grundschulalter von Kindern umfasst und dazu beiträgt, mit präzisen Antworten zur Stabilität der Konstrukte den theoretischen Kenntnisstand zu ergänzen. Die vierte Fragestellung resultierte aus der in der theoretischen Fachdebatte kontrovers diskutierten Frage nach der Stärke und der Richtung der Beeinflussung der Konstrukte kognitive Fähigkeiten und mathematische Leistungen. Die Auffassungen reichen von der vor allem durch kognitive Merkmale beeinflussten mathematischen Schulleistung, bis zur gegenteiligen Auffassung, dass vor allem (Vor-)Wissensaspekte kognitive Merkmale beeinflussen. Daher lautete in diesem Zusammenhang die vierte Frage, inwiefern Fähigkeiten eines Konstrukts die Leistungen des anderen Konstrukts zu einem späteren Zeitpunkt bestimmen und über die gesamte Grundschulzeit Wirkungen erzielen.

Zur Klärung der Fragen wurden im jährlichen Abstand zu vier Messzeitpunkten Kognitive-Fähigkeitstests der KFT-Testverfahrensreihe von Heller und Perleth (2000), Perleth und Heller (2008) und Abou-Koura und Perleth (2005) einerseits und mathematische Schulleistungen andererseits (Deutscher Mathematiktest für erste, zweite, dritte und vierte Klassen) der DEMAT-Testverfahrensreihe von Krajewski, Küspert und Schneider (2002), Krajewski, Liehm und Schneider (2004) sowie Roick, Gölitze und Hasselhorn (2004, 2006) erhoben.

Die Stichprobe basierte auf einer regionalen Totalerhebung aus der Mecklenburger Längsschnittstudie der Jahre 2006 bis 2009 mit einer Beteiligung von 1726 Grundschulkindern der ersten bis vierten Klassenstufen.

Für die Analysen konnten Strukturgleichungsmodelle spezifiziert werden, mithilfe derer es möglich war, Strukturen auch auf latenter Ebene zu ermitteln und Zusammenhänge zuverlässig darzustellen, da sie messfehlerbereinigte komplexe Analysen auf latenter Ebene erlauben und daher in besonderem Maße für präzise Aussagen geeignet sind.

Die Frage der Unabhängigkeit der Konstrukte kognitiver und mathematischer Leistungen (Frage 1) wurde mit Zweifaktoren-Korrelationsmodellen auf latenter Ebene zu den jeweils gleichen Messzeitpunkten spezifiziert. Dabei wurden die Korrelationen der beteiligten Konstrukte bis auf Subtestebene auf der latenten Ebene zum selben Zeitpunkt erfasst. Die Analysen erfolgten unter Berücksichtigung der Mehrebenenstruktur der Daten, sodass unterschiedliche Konstellationen auf Schulebene kontrolliert werden konnten.

Für die Invarianzüberprüfung (Fragestellung 2) wurde ein konstruktspezifisches Latent-State-Modell spezifiziert, während die Überprüfung der zeitlichen Stabilität (Fragestellung 3) mit konstruktspezifischen autoregressiven Strukturgleichungsmodellen erfolgte, die anschließend in ein gemeinsames latentes Variablenmodell mit kreuzverzögerten Effekten (Fragestellung 4) zusammengeführt wurden. Dieses Modell konnte Aufschluss über die gegenseitige Beeinflussung kognitiver Fähigkeiten und mathematische Leistungen über die Zeit geben.

Die Analysen ergaben relevante Ergebnisse, die im Folgenden zusammengefasst aufgezeigt werden:

Die erste Fragestellung sollte einen Beitrag zu den bislang kontrovers diskutierten theoretischen Annahmen leisten, ob die Messung der Konstrukte kognitive Fähigkeiten auf der einen Seite und mathematische Leistungen auf der anderen Seite womöglich identische Fähigkeiten erfassen, bzw. ob vielmehr beide Konstrukte unabhängige Fähigkeiten erfassen. Ob kognitive Fähigkeitstestverfahren und Messungen mathematischer Leistungen möglicherweise identische Konstrukte erfassen oder ob die Unabhängigkeit beider Konstrukte angenommen werden kann, besitzt grundlegende theoretische Bedeutung.

Es wurde folgende Fragestellung formuliert: Inwiefern handelt es sich bei den Konstrukten kognitive und mathematische Leistung um präzise trennbare, unabhängige Bereiche? Zur Klärung der Unabhängigkeit der Konstrukte kognitiver und mathematischer Leistungen wurden Zweifaktoren-Korrelationsmodelle auf latenter Ebene spezifiziert.

Die Ergebnisse dieses konzeptionellen Teils der Arbeit zeigen unter Berücksichtigung der Mehrebenenstruktur der Daten in der Gesamtbetrachtung eine eindeutige faktorielle Differenziertheit der Konstrukte.

Im Detail waren die Korrelationen für die manifesten Variablen zum jeweiligen Subtest des gleichen Konstrukts höher als zum anderen Konstrukt. Gleichzeitig aber lassen sich jedoch auch jeweils die Zusammenhänge zwischen kognitiven Fähigkeiten und mathematischen Schulleistungen erkennen. Auf der latenten Ebene ist interessant zu sehen, dass auch die Korrelationen zwischen den latenten Variablen der jeweils anderen Konstrukte hoch sind und damit konstant über die Zeit bestehen. Eine hohe Korrelation der latenten State Variablen spricht für eine hohe Stabilität interindividueller Unterschiede (Geiser, 2010). Die mittlere Korrelation der latenten Variablen zu den jeweiligen Messzeitpunkten zu dem jeweils anderen Konstrukt bedeutet, dass auf der nicht beobachtbaren Ebene die gleichen Einflüsse wirksam sind.

Es kann insgesamt begründet geschlussfolgert werden, dass die Unabhängigkeit von kognitiven Fähigkeiten und mathematischen Leistungen sowohl auf manifester als auch auf latenter Ebene empirisch belegbar ist.

Die zweite Fragestellung verfolgte das Ziel, die Vergleichbarkeitsfrage der Testverfahren über vier Jahre hinsichtlich der Konstruktvalidität kognitiver Fähigkeiten und mathematischer Leistungen zu klären.

Die Resultate der Invarianzprüfung zeigen folgendes Bild: Die Konstrukte kognitive Fähigkeiten und mathematische Leistungen erwiesen sich über die Grundschulzeit hinweg als invariant, wobei sich die Invarianz für die kognitiven Fähigkeitstests konsistenter nachweisen ließ als für die Testverfahren der Mathematikleistungen.

Kognitive Fähigkeiten (erstes Konstrukt) werden folglich mit den Testverfahren von Heller und Perleth (2000), Perleth und Heller (2008) und Abou-Koura und Perleth (2005) von der ersten Klassenstufe bis hin zur vierten Klassenstufe zuverlässig erfasst. Die Testreihe zur Erfassung kognitiver Fähigkeiten im Grundschulalter ist somit gut geeignet, diese Fähigkeiten von Grundschulkindern im Längsschnitt zu ermitteln. Die Konstruktkonsistenz ist besonders zufriedenstellend zwischen verbalen und nonverbalen Subtests des kognitiven Fähigkeitstests für die jeweils folgende Klassenstufe. Die quantitative Subskala zeigt hier Werte zwischen .12 und .45, welche als zufriedenstellend anzusehen sind. Befriedigende Ergebnisse zeigen sich in der Überprüfung der Konstruktkonsistenz der Subtests der kognitiven Testverfahren (verbal, quantitativ und nonverbal) über zwei Klassenstufen hinweg,

während die Ergebnisse erfreulicherweise keinen weiteren Verlust an Aussagekraft in der Betrachtung über drei Klassenstufen hinweg zeigen.

Für die Testreihe zur Erfassung mathematischer Fähigkeiten (zweites Konstrukt) von Krajewski, Küspert und Schneider (2002), Krajewski, Liehm und Schneider (2004) und Roick, Göllitz und Hasselhorn (2004, 2006) kann nur teilweise davon ausgegangen werden, dass Leistungen, die am Ende der vierten Klassenstufe gemessen werden, auch dasselbe Konstrukt erfassen, das unter diesem Namen in der ersten Klassenstufe bzw. in der zweiten Klassenstufe erfasst wurde. Im Detail ist die Konstruktkonsistenz noch am höchsten für den Bereich Arithmetik der dritten zur vierten Klassenstufe und für den Geometriebereich der zweiten zur dritten und von der dritten zur vierten Klassenstufe. Erfreulicherweise zeigt sich auch eine Konstruktkonsistenz des Geometriebereichs von der zweiten bis zur vierten Klassenstufe. Für die Sachrechenaufgaben können keine Konstruktkonsistenzen nachgewiesen werden, sodass insgesamt für die Testreihe zur Erfassung mathematischer Fähigkeiten über vier Jahre nur partielle Invarianz nachgewiesen werden konnte.

Im Mittelpunkt der dritten Fragestellung stand die Überprüfung der zeitlichen Stabilität interindividueller Unterschiede kognitiver Fähigkeiten und mathematischer Schulleistungen. Sie lassen sich wie folgt zusammenfassen:

In der Gesamtbetrachtung kann formuliert werden, dass Intelligenzunterschiede und Mathematikleistungen vom ersten Grundschuljahr an bis zum Ende der Grundschulzeit (ca. zehnten Lebensjahr) als stabil anzusehen sind.

In der Betrachtung des einzelnen Konstrukts kognitive Fähigkeiten zeigt sich folgendes Bild: Hinsichtlich der Stabilität kognitiver Fähigkeiten weisen die autoregressiven Pfade über ein bis drei Jahre (Pfade erster, zweiter und dritter Ordnung) sowie die Residualvarianzen eine hohe Stabilität der interindividuellen Merkmalsausprägungen auf.

Detaillierter betrachtet kann festgestellt werden, dass Vorhersagen kognitiver Fähigkeiten auf Werte der jeweiligen Folgejahre recht zuverlässig möglich sind. Prognosen aus Leistungen in der ersten Klassenstufe auf Werte in der zweiten Klassenstufe gelingen beispielsweise mit .90. Nur gering dagegen lassen sich wiederum kognitive Fähigkeiten der dritten Klassenstufe aus den Fähigkeiten der zweiten Klassenstufe vorhersagen. Etwas höher, aber dennoch gering sind die Stabilitätseffekte zwischen kognitiven Fähigkeiten der dritten und der vierten Klassenstufe.

Sollen jedoch Stabilitätswerte über zwei Jahre geschätzt werden, so gelingt diese Vorhersage mit Werten von .87 aus der ersten Klassenstufe auf die dritte Klassenstufe sehr

gut, während kognitive Fähigkeiten der vierten Klassenstufe aus den Werten der zweiten Klassenstufe mit geringeren Stabilitätseffekten vorhergesagt werden können.

Die interindividuelle Stabilität kognitiver Fähigkeiten unterliegt damit von der ersten Klassenstufe bis zum Ende der Grundschulzeit lediglich geringen Schwankungen. Die Aussagen basieren vor allem auf den hohen Werten zwischen der ersten und zweiten Klassenstufe, als auch auf den hohen Stabilitätswerten zwischen der ersten und dritten Klassenstufe und mittleren Werten zwischen der zweiten und der vierten Klassenstufe sowie den geringen, aber nachweisbaren Stabilitätswerten zwischen der dritten und der vierten Klassenstufe.

Damit kann geschlussfolgert werden, dass die prognostische Validität kognitiver Leistungen im betrachteten Zeitraum von vier Jahren, insgesamt betrachtet, hoch ist. Werden nunmehr die Stabilitätseffekte mathematischer Leistungen (zweites Konstrukt) im Laufe des Grundschulalters untersucht, zeigen sich folgende Ergebnisse:

Die Spezifikation der autoregressiven Pfade der mathematischen Fähigkeiten zeigt, dass sich mathematische Leistungen in Folgejahren ebenso recht zuverlässig aus früher erbrachten Leistungen vorhersagen lassen. Die Stabilitätseffekte, mit denen sich die Leistungen aus den Vorjahresleistungen vorhersagen lassen, liegen insgesamt zwischen .15 und .88.

In der differenzierten Betrachtung wurde zunächst die Stärke der Beeinflussung auf das jeweilige Folgejahr untersucht. Sie war zwischen der ersten und zweiten Klassenstufe liegt bei .90, zwischen der zweiten und dritten Klassenstufe (aufgrund der angesprochenen Testadaption) ist sie geringer, jedoch nachweisbar. Zwischen der dritten und vierten Klassenstufe sind die Werte erneut im mittleren Bereich. Bei der Überprüfung, inwieweit sich Mathematikleistungen über zwei Jahre vorhersagen lassen, zeigen sich sehr hohe Stabilitätseffekte von .81 zwischen der ersten und dritten Klassenstufe und geringere Werte zwischen der zweiten und vierten Klassenstufe. Über drei Jahre betrachtet lassen sich Mathematikleistungen der vierten Klassenstufe sehr zuverlässig bereits aus den Leistungen in der ersten Klassenstufe vorhersagen (.58).

Insgesamt gesehen weisen die Stabilitätswerte der Schulleistungstests für mathematische Fähigkeiten der ersten bis vierten Klassenstufe sehr zufriedenstellende Resultate auf. Zusammenfassend kann hinsichtlich der Stabilitätsfrage beider Konstrukte festgestellt werden, dass eine hohe interindividuelle Stabilität sowohl kognitiver als auch mathematischer Leistungen im Grundschulalter bestätigt werden kann.

Im Rahmen der vierten Fragefragestellung sollte beantwortet werden, inwiefern vor allem kognitive Fähigkeiten bestimmend für mathematische Leistungen der Folgejahre sind. Zugleich sollte geklärt werden, ob Beeinflussungen in umgekehrter Richtung vorhanden sind, sodass mathematische Leistungen wiederum die kognitiven Fähigkeiten der Folgejahre bestimmen, bzw. inwieweit sich kognitive und mathematische Leistungen gegenseitig beeinflussen. Dabei zeichnet sich diese Arbeit dadurch aus, dass die Ergebnisse nicht nur auf korrelativer manifester Ebene erfasst worden sind, sondern dass es gelungen ist, die Richtung der Beeinflussungen offenzulegen. Die Ergebnisse der Untersuchungen zeigen in der Gesamtbetrachtung folgende Resultate:

Die Betrachtung der einseitigen Einflüsse von kognitiven Fähigkeiten auf spätere mathematische Leistungen (zuerst betrachtete Pfadrichtung) zeigt, dass eine Vorhersage von Mathematikleistungen früherer Zeitpunkte in Folgejahren zuverlässig möglich ist. Dabei ist bereits der Einfluss kognitiver Fähigkeiten der ersten Klassenstufe auf die mathematischen Schulleistungen als recht hoch einzuschätzen. Vorhersagen von Mathematikleistungen der dritten Klassenstufe aufgrund von kognitiven Fähigkeiten, die in der zweiten Klassenstufe erzielt wurden, gelingen in mittlerer Höhe mit .29 (.20 in Modell 3 – Crossed Lagged). Ein ähnliches Ergebnis zeigt die Betrachtung der Wirkung von kognitiven Fähigkeiten der dritten Klassenstufe auf Mathematikleistungen der vierten Klassenstufe (.19 in Modell 3 – Kreuzverzögerte Effekte). Kreuzverzögerte Effekte lassen sich sowohl von kognitiven Fähigkeiten auf mathematische Leistungen als auch und (sehr deutlich) von mathematischen Leistungen auf kognitive Fähigkeiten in Folgejahren nachweisen.

Der Nachweis recht hoher Einflüsse kognitiver Fähigkeiten auf spätere Mathematikleistungen fügt sich übereinstimmend in etablierte Theorien ein. Theoretische Bedeutung haben diese Resultate jedoch dahingehend, dass sich die Wirkungen nicht nur auf die direkten Folgejahre zwischen zwei Messzeitpunkten, sondern über mehrere Messzeitpunkte hinweg, über die gesamte Grundschulzeit, eindeutig nachweisen lassen.

Die im Anschluss betrachteten Pfade der Gegenrichtung zeigen die Wirkung von mathematischen Fähigkeiten auf die kognitiven Fähigkeiten in Folgejahren. Die Gesamtbetrachtung zeigt die Effekte früherer mathematischer Leistungen auf spätere kognitive Fähigkeiten in deutlicher Höhe. Im Detail zeigt die Prognosekraft von Mathematikleistungen der ersten Klassenstufe auf kognitive Fähigkeiten der zweiten Klassenstufe eine Ausprägung in mittlerer Höhe von .34. Die Betrachtung von Mathematikleistungen der zweiten Klassenstufe auf kognitive Fähigkeiten der dritten Klassenstufe wiederum zeigt eine etwas höhere Vorhersagekraft von .38, während Mathematikleistungen der dritten Klassenstufe

die kognitiven Fähigkeiten der vierten Klassenstufe mit .63 am höchsten vorherzusagen vermögen. Der Nachweis von Effekten früherer Mathematikleistungen auf kognitive Fähigkeiten (Gegenrichtungspfade) der direkten Folgejahre hingegen ist über den betrachteten Vierjahreszeitraum deutlich und zeigt überdies zunehmende Wirkung bis zu einer Höhe von .63 zum Ende der Grundschulzeit. Damit sind die Wirkungen von Mathematikleistungen auf kognitiven Fähigkeiten stärker als die Wirkungen von kognitiven Fähigkeiten auf Mathematikleistungen (.29 in Modell 2 und im Crossed-Lagged-Modell 3 mit .19).

Insgesamt weisen diese Pfadrichtungsanalysen deutliche Wirkungen von Mathematikleistungen auf kognitive Leistungen der Folgejahre auf. Zusammenfassend lässt sich hinsichtlich der Frage der gegenseitigen Beeinflussung feststellen: Die gleichzeitige Überprüfung des Verlaufs der Konstrukte kognitive Fähigkeiten und mathematische Fähigkeiten anhand der statistisch bedeutsamen Residualvarianzen aus den Stabilitätsanalysen über die Zeit zeigt eine zunehmende Prognosekraft der kreuzverzögerten Effekte über die Jahre. Im Rahmen der theoretischen Debatten ist damit auch eine empirische Unterstützung für die Investitionstheorie (Gf-Gc) von Cattell gelungen, die aufgrund ihrer theoretischen, gut begründbaren (letztlich nur empirisch nicht belegten) Annahmen als theoretische Basis für die vorliegende Arbeit angenommen wurde. Bisherige Versuche zum Beispiel von Watkins, (2007), die Gegenseitigkeit der Beeinflussung beider Konstrukte empirisch zu belegen, waren ernüchternd verlaufen. Mit der vorliegenden Arbeit wird Cattells Annahme, dass erworbene kristallisierte Intelligenz, die im Zusammenhang mit dem akademisch bzw. schulisch erworbenen Bildungswissen steht, auf die fluiden Intelligenzanteile rückwirkt und diese positiv beeinflusst, erstmals empirisch auf latenter Ebene über vier Messzeitpunkte für das Grundschulalter nachgewiesen. Die Annahme, dass schulische Bildung die schulische Leistung bestimmt, ist mit der Integration der Position, dass kognitive Fähigkeiten die Bildungsleistungen beeinflussen, zu vereinen. Denn die Fähigkeit zu lernen (im Sinne des kognitiven Potentials), die mittels fluider Intelligenztests gemessen wird, kann sich in hohen Fähigkeiten widerspiegeln. Diese hohe Fähigkeit wurde häufig als einziger Prädiktor späterer akademischer Leistung angenommen. Wenn also nur die fluiden Intelligenzanteile die abschließende akademische Leistung vorhersagen dürfen, so sollte die prädiktive Gültigkeit des Einflusses fluider Intelligenz bestätigt werden. Aber wenn nun auch von der akademischen Leistung als Prädiktor ausgegangen wird, wird der ausnahmslose Beitrag der flüssigen Intelligenz reduziert, da sich akademische Leistung zum späteren Zeitpunkt ursächlich aus der akademischen Leistungsmessung zum vorherigen Zeitpunkt prognostizieren lässt und damit repräsentiert wird. Damit wird auch die Position integriert, nach der schulische

Leistung auch durch die anfänglichen schulischen Leistungen vermittelt wird und die Annahme, dass wiederum die schulischen Leistungen auch die kognitiven Fähigkeiten zu späteren Zeitpunkten bestimmen.

Mit den Ergebnissen der vorliegenden Arbeit ist es also empirisch gelungen, die wechselseitigen zeitlichen Einflüsse fluider und kristallisierter Intelligenzanteile durch die wechselseitigen Einflüsse kognitiver Fähigkeiten, die wiederum in verbale, nonverbale und quantitative Bereiche differenziert wurden, darzustellen. Die Ergebnisse sind anschlussfähig an das von Perleth entwickelte Münchner Begabungsmodell (Heller, 1992c und Perleth & Heller, 1994) sowie an dessen Erweiterung (Perleth, 1997), da dieses explizit die Vorwissensanteile berücksichtigt.

4.2 Ergebnisse 2 – Invarianz – KFT

Um im Folgenden die Invarianz des Testverfahrens für die kognitiven Fähigkeitsskalen zu überprüfen, wird zunächst ein schwach restriktives Invarianzmodell spezifiziert. Unter den Bedingungen der Gleichsetzung nach Widaman und Reise (1997) der Intercepts des 1. Indikators (KFT quantitativ) und der Schätzung der Intercepts des 2. und 3. Indikators sollen die Erwartungswerte der latenten Variablen frei geschätzt werden. Nach Geiser (2011) spricht eine hohe Korrelation der latenten State-Variablen für eine hohe Stabilität interindividueller Unterschiede. Für den KFT wurden die Daten von 1665 Fällen berücksichtigt.

4.2.1 Konfigurale Invarianz – KFT

Die folgende Abbildung zeigt die Korrelation zwischen den latenten Variablen im Modell mit konfiguraler Invarianz des kognitiven Fähigkeitstests.

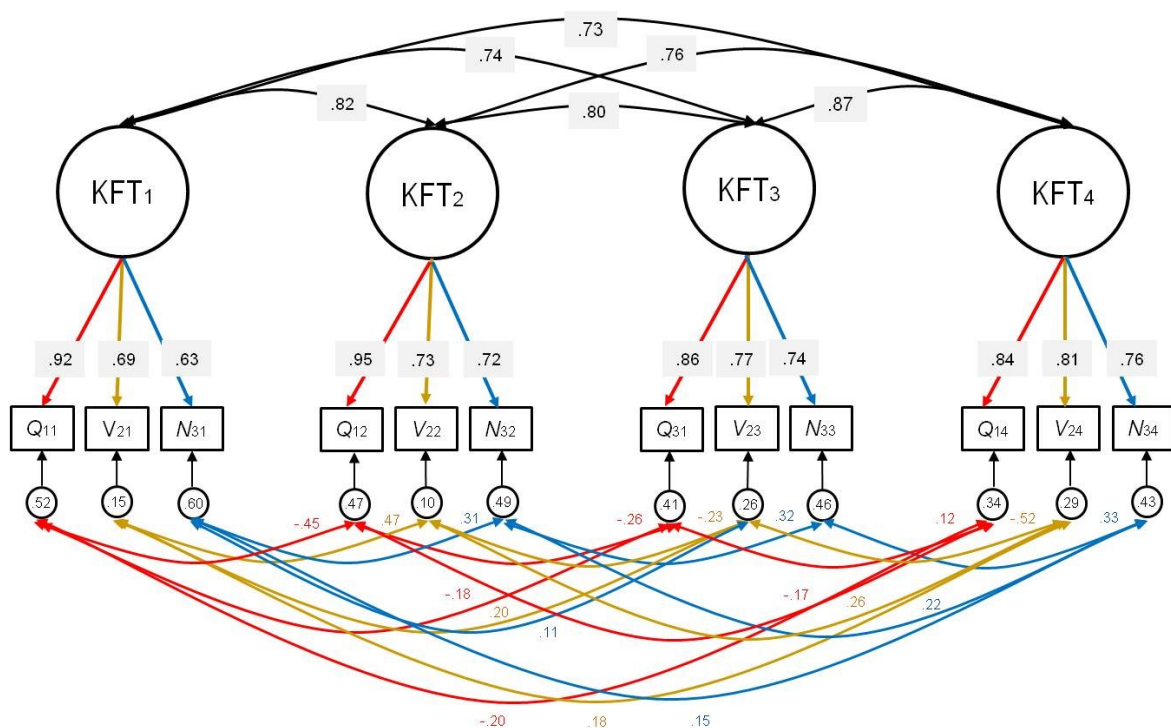


Abbildung 35 Strukturmodell zur Testung der Invarianz;
V = verbal; Q = quantitativ; N = nonverbal.

Die Residualkorrelationen für den verbalen Untertest des KFT 1 und KFT 2 sind mit .47 im mittleren Bereich, während sie zwischen dem KFT 1 und dem KFT 2 deutlich niedriger sind. Zwischen dem KFT 2 und dem KFT 3 zeigen sich etwas höhere Korrelationen in Höhe

von .23 und .26 zwischen dem KFT 2 und dem KFT 3. Die Korrelationen zwischen dem KFT 1 und dem KFT 4 weisen wiederum eine mittlere Höhe von .52 auf, während die standardisierten Residualkorrelationen Werte im leicht negativen Bereich aufweisen. Lediglich zwischen dem KFT 3 quantitativ und dem KFT 4 quantitativ weisen sie einen Wert von .12 auf.

Tabelle 21

Effektstärken und Faktor reliabilitäten (KFT) für vier Messzeitpunkte

	λ^2	ω
Messzeitpunkt 1		
KFT 1 verbal	.48	
KFT 1 quantitativ	.85	.80
KFT 1 nonverbal	.40	
Messzeitpunkt 2		
KFT 2 verbal	.53	
KFT 2 quantitativ	.90	.84
KFT 2 nonverbal	.51	
Messzeitpunkt 3		
KFT 3 verbal	.59	
KFT 3 quantitativ	.78	.83
KFT 3 nonverbal	.52	
Messzeitpunkt 4		
KFT 4 verbal	.66	
KFT 4 quantitativ	.75	.85
KFT 4 nonverbal	.56	

λ^2 = quadrierte Faktorladung; ω = Faktor reliabilität.

In der Betrachtung der standardisierten Residualkorrelationen zwischen KFT 1 nonverbal und KFT 2 nonverbal zeigt sich eine Höhe von .31, während die Korrelation zwischen KFT 1 und KFT 3 eine Höhe von .11 aufweist. Die Residualkorrelation zwischen der nonverbalen Subskalen des KFT 1 und dem KFT 4 zeigt eine Höhe von .15. Die korrelativen Zusammenhänge zwischen KFT 2 und KFT 3 in den nonverbalen Subskalen fallen mit .32 leicht höher aus, während der KFT 2 nonverbal mit dem KFT 4 nonverbal einen Wert von .21 aufweist. Die Korrelationen auf Residualebene sind zwischen KFT 3 und KFT 4 mit .33 noch am höchsten.

Um Aussagen zur Reliabilität der Faktoren zu ermitteln, wurde die Maximum-Likelihood-Methode gerechnet. Die Diskrepanz zwischen den vom Modell implizierten Zusammenhängen (Kovarianzen) und den beobachteten Zusammenhängen sollte minimal sein. Um das Maximum-Likelihood-Verfahren anzuwenden, sind zuvor die Voraussetzungen zu prüfen. So sollte die Stichprobe hinreichend groß sein, die Beobachtungen unabhängig und die Daten multivariat normalverteilt sein.

Bei kleineren Verletzungen ist das Verfahren recht robust und rechtfertigt eine Anwendung (Eid et al., 2013). Für die vorliegenden Daten sind diese Bedingungen erfüllt (siehe Abschnitt zur Normalverteilungsprüfung).

Tabelle 18 gibt Auskunft über die Effektstärken und die Faktorreliabilitäten. Die Faktorladungsstruktur zeigt durchweg akzeptable Werte, die jeweils am höchsten für die Subskalen der Arithmetik aller vier Verfahren ausgeprägt sind.

Die Faktorreliabilität gibt Auskunft über die Effektstärke und ist ein anschauliches Maß, wie viel Prozent der Gesamtvarianz über die systematische Varianz erklärt werden kann. Die quadrierte Faktorladung entspricht der Summe aus systematischer Varianz und Residualvarianz und kann Werte zwischen 0 und 1 annehmen, während die maximal erklärte Gesamtvarianz 1 betragen kann, (Eid et al., 2013).

Mit Werten über die jeweiligen Subskalen der KFT-Testformen zwischen .78 im 1. Messzeitpunkt und .84 im 2. Messzeitpunkt, .83 zum 3. Messzeitpunkt und .85 ist die Effektstärke im sehr zufriedenstellenden Bereich. Dies bedeutet, dass die Kinder über die vier Jahre hinweg größtenteils äquivalente Werte erzielten. Die quadrierte Faktorladung des KFT gibt Auskunft über die Reliabilität der Einzelskalen. Sie erklärt die durch die Modellspezifikation erzielte Aufklärung der State-Faktor-Varianz und kann als Prozentwert gelesen werden. Sie weist in der ersten Klassenstufe nur eine knapp durchschnittliche Reliabilität in den Skalen „verbal“ und „nonverbal“ auf. Der Wert von .85 zeigt eine durch die State-Varianz eine Rückführbarkeit auf die KFT-Dimensionen zu 85 Prozent. Die restlichen

15 Prozent der State-Varianz gehen auf Messfehler zurück. Die durch Messgelegenheiten am wenigsten beeinträchtigte Subskala ist die der mathematischen Fähigkeiten. Die quantitative Skala besitzt im 2. Testjahr eine Reliabilität von .90 und hat damit einen äußerst hohen erklärten Anteil an wahren, von Messeinflüssen wenig beeinträchtigten mathematischen Fähigkeiten. Das ist für den quantitativen Teil des gesamten Testverfahrens über die Testformen hinweg die höchste Reliabilität. In der dritten und vierten Klassenstufe liegen die Werte der quadrierten Faktorladung jedoch ebenso im oberen Bereich von .78 und .75. Auch die Subskalen „verbale Fähigkeiten“ der Messzeitpunkte 1, 2, 3, 4 zeigen einen weiteren Anstieg der Reliabilität von .48 über .53 und .59 auf bis zu .66. Für die nonverbalen Skalen kann erfreulicherweise ein ebensolcher Anstieg der Reliabilitätswerte verzeichnet werden von .48 zum 1. Messzeitpunkt, zu .59 im 2. Messzeitpunkt und .59 zum 3. Messzeitpunkt auf .66 im 4. Messzeitpunkt.

Diskussion und Interpretation

Ein Teil der Residualkorrelationen sind jedoch bedeutsam negativ. So ist vor allem der nichtaufklärbare Anteil zwischen dem quantitativen Teil des KFT 1 und den quantitativen Residualvarianzen des KFT 2, des KFT 3 und des KFT 4 negativ korreliert. Ebenso findet sich eine negative Residualkorrelation zwischen dem quantitativen Teil des KFT 2 und den quantitativen Bereichen des KFT 3 und des KFT 4.

Dieses Ergebnis könnte sich inhaltlich begründen lassen. So sind die KFT-Aufgaben des ersten Messzeitpunktes von den Kindern stark mit dem Hörverstehen assoziiert. Die gleichzeitige Belastung des Arbeitsspeichers durch das Verarbeiten auditiver Informationen und deren Relevanzprüfung zur Problemlösung kann diese Ergebnisse erklären. Die Kinder der ersten Klassenstufe können die Aufgaben zumeist noch nicht selbstständig lesend erfassen, deshalb werden ihnen die Aufgaben vom Testleiter vorgelesen. Somit ist zu den Fähigkeiten, die Aufgaben auf kognitiver Fähigkeitsebene lösen zu können, auch der Arbeitsspeicher stark gefordert. In den nachfolgenden Jahren nimmt diese Arbeitsspeicherbelastung durch die Replizierbarkeit der Aufgaben ab, denn die Kinder haben dann die Möglichkeit, Aufgaben selbstständig erneut zu lesen. Auch die Modifikationsvorschläge, die sich im Output zeigen, um die Modellpassung zu verbessern, gehen in diese Richtung. Sie sehen vor, dass quantitative Aufgabenanteile mit verbalen Anteilen des KFT korrelieren können. Ebenso zeigen die Modifikationsvorschläge, dass nonverbale Anteile mit verbalen Aufgaben des KFT 1 korrelieren wollen.

Die Faktorladungen des KFT 1 bis 4 mit den Subskalen verbal, quantitativ und nonverbal weisen auf latenter Ebene durchweg hohe bis sehr Korrelationen auf, was als sehr zufriedenstellend interpretiert werden kann.

Da die Passung noch nicht sehr zufriedenstellend war (hoher χ^2), wurde zunächst ein Modellfehler angenommen und den Handlungsempfehlungen für solche Fälle gefolgt. Die erste Option ist eine Kreuzvalidierung mit je der Hälfte der Stichprobe. Um jedoch modell-spezifische Konzeptionsfehler sicher ausschließen zu können und die oben angenommene inhaltliche Begründung statistisch annehmen zu können, wurden zwei Methoden der Kreuzvalidierung der Stichprobe vorgenommen. Zum einen erfolgte die Differenzierung der Stichprobe nach Geschlecht. Auch hier zeigten sich die gleichen Effekte über die Geschlechter hinweg, sodass für die negative Korrelation der Messfehler inhaltliche Gründe angenommen werden müssen, die sich über die, wenn auch geringe, Methodendifferenz zwischen KFT 1-2 und KFT 3 und KFT 4 (Autorenwechsel) aufklären lassen.

Die zweite Option ist die Ziehung zweier Zufallsstichproben in SPSS, um das Modell unter den veränderten Bedingungen zu spezifizieren. Wenn inhaltliche Gründe vorliegen, sollte sich das gleiche Ergebnis zeigen. Auch die Ziehung einer Zufallsstichprobe zeigte inhaltlich gleiche Ergebnisse, sodass die Argumentation auf inhaltliche Gründe schlüssig ist.

Die negative Korrelation zeigte sich zudem nur auf latenter Ebene, nicht aber auf manifester Ebene, sodass auf latenter Ebene mit der Stichprobengröße argumentiert werden kann, denn die Residualvarianz ist zusätzlich sehr eingeschränkt, da der Standardfehler bei der Stichprobengröße relativ hoch ist.

Der nicht aufgeklärte Anteil der Residualvarianz kann daher ebenso auf die Methodenunterschiedlichkeit zurückgeführt werden.

Die Korrelationen zwischen den latenten Variablen des KFT 1 und KFT 2 sind mit .82 sehr hoch und zwischen dem KFT 1 und KFT 2 mit .74 ebenfalls recht hoch. In fast gleicher Höhe weist die Korrelationen zwischen den latenten Variablen des KFT 1 und KFT 2 einen Wert von .73 auf. Die Korrelationen zwischen dem KFT 2 und dem KFT 3 sind wiederum mit .81 sehr hoch, und fast ebenso hoch sind die Korrelationen zwischen dem KFT 2 und dem KFT 4 mit .76. Die Korrelationen zwischen den latenten Variablen des KFT 3 und KFT 4 sind mit einem Wert in der Höhe von .87 am höchsten. Um im Zuge der Invarianzprüfung einen stärkeren Grad der Messinvarianz zu testen und damit mehr Zusammenhänge aufklären zu können, wird im folgenden Schritt ein Modell mit metrischer Invarianz getestet.

4.2.2 Metrische Invarianz – KFT

Ein etwas stärkerer Grad der Messeinvarianz ist gegeben, wenn Faktorladungen über die Zeit konstant bleiben und somit schwache faktorielle (metrische) Invarianz erreicht werden kann.

Für die entsprechende Modellspezifikation werden die Intercepts des 1. Indikators (KFT quantitativ) auf 0 gesetzt und die Indikatoren des verbalen Teils des KFT fixiert sowie die nonverbalen Skalen des KFT ebenfalls gleichgesetzt. Es zeigt sich eine Modellverschlechterung. Der χ^2 des Modells weist einen Wert von 474.963 bei 36 Freiheitsgraden auf. Der CFI ist akzeptabel, der RMSEA relativ hoch. Auch auf dieser Restriktionsebene zeigen sich die bereits erwähnten Residualkorrelationen in den quantitativen Aufgabenanteilen des 1. Messzeitpunktes auf latenter Ebene, die sich jedoch insofern verbessert haben, als dass sie sich zum zweiten und dritten Messzeitpunkt nicht mehr im signifikanten Bereich befinden, lediglich zwischen KFT 1 (quantitativ) und KFT 4 (quantitativ) liegen sie im Signifikanzbereich.

4.2.3 Skalare Invarianz – KFT

Von starker faktorieller Invarianz kann ausgegangen werden, wenn darüber hinaus auch die Ladungen und Intercepts über die Zeit invariant bleiben und somit skalare Invarianz erreicht werden kann. Es erfolgt eine Restriktion des verbalen Aufgabenbereichs über die Messzeitpunkte auf 1 sowie eine Restriktion der nonverbalen Aufgabenbereiche auf 1. Die Fixierung der Intercepts des ersten Indikators erfolgt auf 0, wobei die Intercepts des zweiten und dritten Indikators auf 1 gesetzt werden. Die Schätzung der Erwartungswerte der latenten Variablen zeigt folgendes Bild:

Die Modellpassung unter diesen Bedingungen zeigt einen χ^2 von 4398.268 bei 42 Freiheitsgraden und einem CFI von .53 und einem RMSEA von .25.

Tabelle 22

Modellvergleich Invarianz KFT, Passung der einzelnen Modelle zueinander

	Modell unter konfiguraler IA	Modell unter metrischer IA mit Residualkovarianzen	Modell unter skalarer IA
χ^2	153.271	474.963	4398.268
<i>df</i>	30	36	42
<i>p</i>	0.000	0.000	0.000
<i>CFI</i>	.99	.93	.53
<i>RMSEA</i>	.05	.09	.25

IA = Invarianzannahme; χ^2 = Diskrepanz; *df* = Freiheitsgrade; *p* = Signifikanz; *CFI* = Comparative Fit Index; *RMSEA* = Root Mean Square Error of Approximation.

Der Modellvergleich mit dem schwächeren Modell zeigt keine schlechtere Passung des Modells auf skalarer Ebene. Ein Blick auf die Korrelation der Residualvariablen zeigt jedoch mehrere Korrelationen im negativen Bereich. Der χ^2 -Test der Modellfitness unter konfiguraler Invarianz zeigt einen Wert von 153.271 und 30 Freiheitsgrade. Der RMSEA besitzt einen Wert von .05 und der CFI ist mit .99 als sehr gut zu bezeichnen. Die nachfolgende Tabelle zeigt die Passungswerte der Modelle unter den verschiedenen Restriktionen.

4.2.4 Invarianz Modellvergleich – KFT

Der Modellvergleich der für die Prüfung der Messinvarianz des kognitiven Testverfahrens eingesetzten Modelle kann anhand des χ^2 -Differenzentests zeigen, dass das Modell unter metrischer Invarianzbedingungen eine schlechtere Passung aufwies, ebenso wie das Modell zur Testung skalarer Invarianz als das Modell unter konfiguralen Invarianzbedingungen.

Tabelle 23

Modellvergleich χ^2 -Differenzentest KFT, Beurteilung der Passung der Modelle zueinander

	Modell unter metrischer Invarianzannahme mit Residualkovarianzen	Modell unter skalarer Invarianzannahme mit Residualkovarianzen
	vs.	vs.
	Modell unter konfiguraler Invarianzannahme mit Residualkovarianzen	Modell unter metrischer Invarianzannahme mit Residualkovarianzen
$\Delta\chi^2$	3210.69	39230.31
Δdf	6	6
$p (\chi^2\text{-Test})$.00	.00

$\Delta\chi^2$ = Differenz Diskrepanz; Δdf = Differenz Freiheitsgrade; p = Signifikanz.

Die Kontrolle der Mehrebenenstruktur berücksichtigt die Clusterung der Daten in verschiedenen Schulen. Wie die folgende Tabelle zeigt, sind die Unterschiede jedoch nicht bedeutsam.

Tabelle 24

Passungswerte zur Prüfung der Invarianz (jeweils ohne und mit Kontrolle der Mehrebenenstruktur)

Modell konfigurale Invarianz	ohne MEK	mit MEK
χ^2	153.27	126.52
<i>df</i>	30	30
<i>CFI</i>	.99	.99
<i>RMSEA</i>	.05	.04
<i>SRMR</i>	.03	.03

χ^2 = Diskrepanz; *df* = Freiheitsgrade; *p* = Signifikanz; *CFI* = Comparative Fit Index; *RMSEA* = Root Mean Square Error of Approximation.

4.3 Ergebnisse 2 – Invarianz DEMAT

Analog zu dem Vorgehen der Prüfung der Invarianz bei den Leistungstests der kognitiven Fähigkeiten soll die Messinvarianz der Testverfahren überprüft werden.

Für den DEMAT wurden die Daten von 1576 Fällen berücksichtigt. Auffällig ist, dass 22,79 % der Kinder die Sachrechenaufgaben des Testverfahrens für die erste Klassenstufe vollständig lösen konnten. Die Testautoren weisen diese aber explizit als „Anwärmaufgaben“ für die Kinder aus, sodass hier diese zu leicht erscheinenden Aufgaben beabsichtigt worden sind (Krajewski et al., 2002). In den folgenden Kapiteln soll Auskunft über die schrittweise vollzogenen Prüfungen der Invarianzbedingungen gegeben werden.

4.3.1 Konfigurale Invarianz – DEMAT

Zunächst wurde ein konfigurales Modell des mathematischen Schulleistungstests spezifiziert. Ferner wurden die Korrelationen zwischen den latenten Variablen zugelassen. Weiterhin wurden die Korrelationen zwischen den Residuen zugelassen, um zu prüfen, in welchen Bereichen Korrelationen vorhanden sind. Wie in der folgenden Tabelle ersichtlich, tragen die Subskalen der Arithmetikbereiche jeweils die höchsten Faktorladungen. Unter den beiden Subskalen Sachrechnen und Geometrie besitzt die Subskala Sachrechnen stabil über alle vier Jahre die höheren Faktorladungen. Die Residualkorrelationen zeigen mittlere

Zusammenhänge, nur vier sind statistisch bedeutsam. So weisen die Faktorladungen in den Messverfahren für die erste und zweite Klassenstufe einen auffällig niedrigen Wert mit .18 bzw. .20 im Geometriebereich aus.

Tabelle 25

Effektstärken und Faktorreliabilitäten (DEMAT) für vier Messzeitpunkte

	λ^2	ω
Messzeitpunkt 1		
DEMAT 1 Geometrie	.18	
DEMAT 1 Arithmetik	.73	.72
DEMAT 1 Sachrechnen	.53	
Messzeitpunkt 2		
DEMAT 2 Geometrie	.20	
DEMAT 2 Arithmetik	.81	.75
DEMAT 2 Sachrechnen	.55	
Messzeitpunkt 3		
DEMAT 3 Geometrie	.39	
DEMAT 3 Arithmetik	.61	.71
DEMAT 3 Sachrechnen	.41	
Messzeitpunkt 4		
DEMAT 4 Geometrie	.37	
DEMAT 4 Arithmetik	.50	.73
DEMAT 4 Sachrechnen	.60	

λ^2 = quadrierte Faktorladung; ω = Faktorreliabilität.

Somit wird deutlich, dass nur ein relativ geringer Erklärungswert der mathematischen Leistungen auf diese Skalen entfällt, während die weiteren Skalen durchaus gute Werte für die

quadrierte Faktorladung von .37 im Geometriebereich des DEMAT 3 bis .81 in der Arithmetiksubskala des DEMAT 2 aufweisen. Die sehr niedrigen Werte von .18 und .20 lassen sich inhaltlich mit der relativ hohen Aufgabenunterschiedlichkeit begründen, sodass das erste Modell der Messinvarianz mit indikatorspezifischem Faktor nicht konvergierte.

Die folgende Abbildung veranschaulicht die Residualkorrelationen (grau markiert) für die Testung der konfiguralen Invarianz.

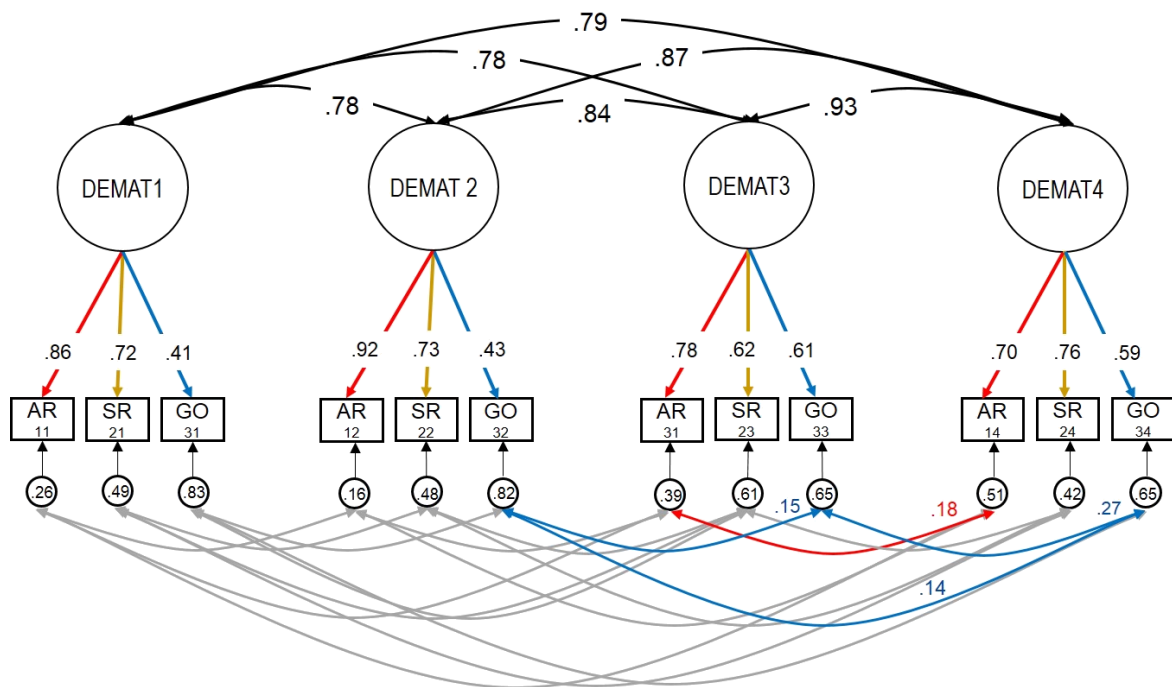


Abbildung 36 Prüfung der Messinvarianz – Stufe I; DEMAT-Residualkorrelationen; AR = Arithmetik; GO = Geometrie; SR = Sachrechnen.

Diese Residualkorrelationen zeigen nur mittlere Zusammenhänge. Daher wurden die Residualkorrelationen in mittlerer Höhe theoriegeleitet nach den Empfehlungen von Sass (2011) aus den Analysen ausgeschlossen, da sich dieses Ergebnis inhaltlich mit der relativ hohen Aufgabenunterschiedlichkeit (unterschiedliche Faktorenstruktur) begründen lässt.

Das Modell der Messinvarianz wird nun mit den verbleibenden vier Residualkovarianzen überprüft, da dieses eine höhere Übereinstimmung mit den Daten aufweist. Die nachfolgende Abbildung veranschaulicht die Ergebnisse.

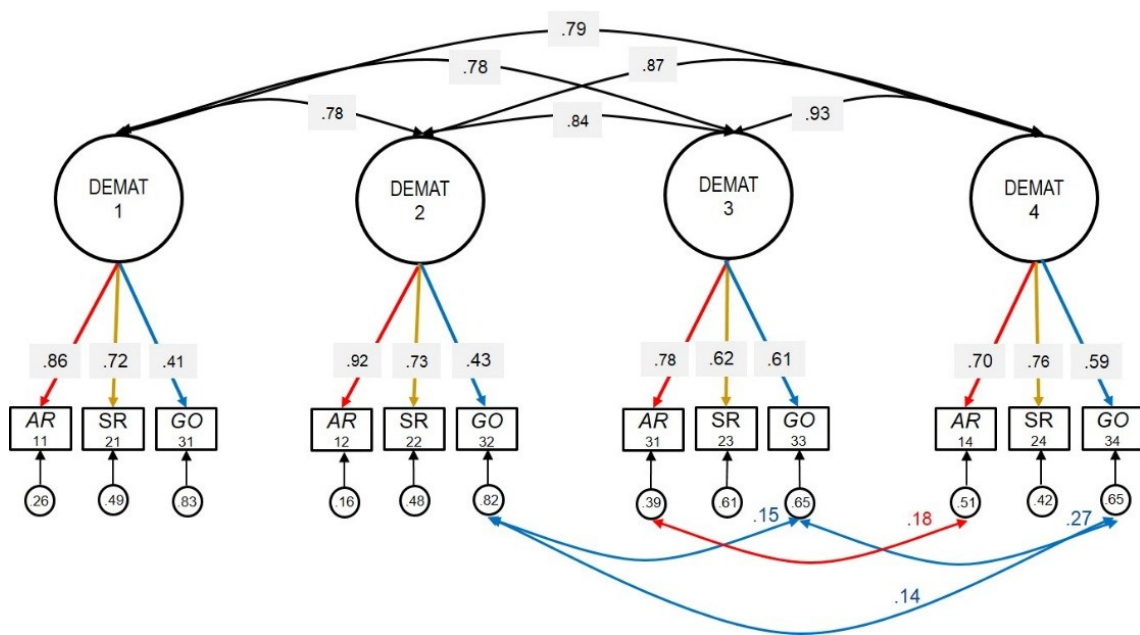


Abbildung 37 Prüfung der Messinvarianz – Stufe II; mit signifikanten DEMAT- Residualkorrelationen; AR = Arithmetik; GO = Geometrie; SR = Sachrechnen.

Die vier zugelassenen Residualkorrelationen sind zwischen der Subskala Geometrie der zweiten und dritten Klassenstufe mit .15 nicht sehr hoch. Ähnliche Ausprägung zeigt die Residualkorrelationen zwischen der Subskala Geometrie der 2. Klasse und der 4. Klassenstufe mit .14. Die Korrelationsbeziehung der Residuen zwischen der Geometrieskala der 3. Klassenstufe und der 4. Klassenstufe ist mit .27 etwas deutlicher ausgeprägt. Die Residualkorrelation zwischen der Arithmetikskala der 3. und der 4. Klassenstufe hat eine niedrigere Ausprägung mit .18.

Die quadrierten Faktorladungen sind relativ gering mit Werten von .17 im Geometriebereich des DEMAT 1 bis zu hohen Werten mit .84 in der Arithmetikskala des DEMAT 2. In diesem konfiguralen Invarianzmodell zeigt sich nunmehr ein sehr guter χ^2 von 76.595 mit 44 Freiheitsgraden und einem fast perfekten CFI von .99 sowie einem sehr gutem RMSEA von .02 und einem SRMR von .02. Damit sind die Zusammenhänge nicht sehr hoch, aber durchaus im moderaten Bereich und können die Zusammenhangsstruktur ausweisen.

Die Korrelationen zwischen den latenten Variablen (standardisiert) in dem Latent-State-Modell mit vier Residualkovarianzen sind in zwischen dem DEMAT 1 und dem DEMAT 2 mit .78 hoch. Das gleiche Ergebnis zeigt sich in der Korrelationsbeziehung zwischen

dem DEMAT 1 und DEMAT 2 und in fast gleicher Höhe auch zwischen DEMAT 1 und DEMAT 3. Noch höhere Korrelationen zeigen sich vom DEMAT 2 zum DEMAT 3 in Höhe von .84 und vom DEMAT 2 zum DEMAT 4 in Höhe von .87, während die höchsten Korrelationen zwischen dem DEMAT 3 und dem DEMAT 4 in Höhe von .93 bestehen.

Auf der Grundlage eines Faktorenmodells höherer Ordnung lässt sich die faktorielle Invarianz über die Grundschulzeit partiell recht gut nachweisen. So zeigen sich die höchsten Skalenzusammenhänge in der Arithmetik über die Testverfahren der 1. bis 4. Klassenstufe hinweg. Diese Faktorladungen zeigen weiter die größte Übereinstimmung in den Aufgaben des Geometriebereichs der 2. und 3. Klassenstufe und des Arithmetikbereichs der 3. und 4. Klassenstufe. Die Erklärungsanteile der Gesamtwerte der jeweiligen Klassenstufentests des DEMAT sind fast identisch mit dem oben berichteten Werten. Zu erkennen ist der leichte Anstieg der Erklärungsanteile der Subtests Geometrie und Sachrechnen innerhalb der Messzeitpunkte. Die Subskalen „Arithmetik“ tragen im DEMAT mit .70 in der vierten Klassenstufe bis .92 Faktorladung zur Erklärung der mathematischen Gesamtfähigkeiten bei. Die Subskalen „Geometrie“ tragen mit einer Ladung von .42 im DEMAT 1 bis .62 im DEMAT 3 ebenfalls recht hohe Anteile. In den Subskalen „Sachrechnen“ betragen die Anteile .64 im DEMAT 3 bis .76 im DEMAT 4. Im folgenden Kapitel sollen nun die Ergebnisse für die Invarianzbedingungen für die Annahme metrischer Invarianz berichtet werden.

4.3.2 Metrische Invarianz – DEMAT

Im folgenden, etwas stärker restriktiven Messinvarianzmodell wird die Faktorladung gleichgesetzt, um so eine metrische Invarianz zu erreichen und das Modell mit metrischer Invarianz getestet. Eine schwache faktorielle Invarianz kann angenommen werden, wenn die Faktorladungen eine Gleichsetzung zulassen und der Modellfit sich in Vergleich zum konfiguralen Modell statistisch nicht bedeutsam verschlechtert. Dazu erfolgt eine Fixierung des 1. Indikators (Subtest Arithmetik) auf 1 sowie die Gleichsetzung der Subskalen des Bereichs Sachrechnen über die Messzeitpunkte sowie auch die Gleichsetzung der Geometrieaufgaben über die Messzeitpunkte. Nachfolgend wird die freie Schätzung der Erwartungswerte der latenten Variablen angefordert. Die Subskala Arithmetik wird weiterhin auf 1 fixiert, da sie die beste Itemstruktur über die Zeit besitzt und auch die höchste Anzahl an Items liefert.

Damit empfiehlt sich, die Items des Geometriebereichs der 1. Klassenstufe wegen der oben genannten, geringen Aussagekraft aus der Modellspezifikation zu entfernen. Diese Aufgabe findet in der Konzeption der Testautoren eine Rolle als Anwärmaufgabe Krajewski

(2002) und daher wurde ihre Konstruktion mit eher geringerer Aussagekraft dennoch zugelassen.

Die Passung des metrischen Invarianzmodells liefert zunächst eine deutlich schlechtere Passung mit einem χ^2 -Test of Model Fit von 609.619 und 50 Freiheitsgraden. Der RMSEA besitzt einen Wert von .08 und einen CFI von .88. Der SRMR beträgt .16 und die quadrierte Faktorladungen liegen in den Bereichen .04 in der Geometrieskala des DEMAT 4 und .78 in der Arithmetikskala des DEMAT 2. Es wird nunmehr versucht, das folgende Modell mit partieller metrischer Invarianz mit 4 Residualkovarianzen zu spezifizieren.

Für die Überprüfung einer partiellen metrischen Invarianz wird die Faktorladung des ersten Indikators (Subskala Arithmetik) weiterhin auf 1 fixiert, es erfolgt die Gleichheitsrestriktion der Subskala Sachrechnen über die Zeit und zudem die Gleichsetzung der Subskala Geometrie über die Messzeitpunkte.

Es zeigt sich eine Verbesserung der Modellpassung. Der χ^2 -Test of Model Fit weist einen Wert von 76.597 auf und der df-Wert kann 45 Freiheitsgraden aufzeigen. Der CFI zeigt sehr gute Indizes von .99. Der RMSEA hat einen guten Wert von .02. Die quadrierte Faktorladung liegt in der Skala Geometrie am niedrigsten mit einem Wert von .17 und am höchsten in der Arithmetikskala des DEMAT 2. Insgesamt liegen sie erfreulicherweise im oberen Bereich.

Eine Kontrolle der Sachrechenaufgabenskonsistenz der DEMAT-Testhefte der Klassenstufe 1 zeigt inhaltlich wenig Ähnlichkeit mit den Sachrechenaufgaben der dritten und vierten Klassenstufe der DEMAT-Verfahren. Das lässt sich leicht mit der starken Lehrplanorientierung der Sachrechenaufgabenkonstruktion in den entsprechenden Klassenstufen begründen. Die Sachrechenaufgabe der Klassenstufe 1 kann daher nicht in die Modellspezifikation einfließen. Die Ähnlichkeitsprüfung der Sachaufgaben der Klassenstufen 3 und 4 zeigt starke inhaltliche Ähnlichkeit auf lediglich schwierigerem Niveau in der Klassenstufe 4. Diese werden daher in der Modellspezifikation fixiert.

Die Gleichsetzung der Arithmetikaufgaben über die Messzeitpunkte gelingt gut, ebenso die Gleichsetzung der Geometrieaufgaben. Die Integration aller Sachaufgaben über die vier Messzeitpunkte ist jedoch, wie bereits oben beschrieben, nicht möglich. Die Modellfitinformationen zeigen sehr gute Passung auf die Daten.

In einem nächsten Schritt soll ein Modell mit starker faktorieller Invarianz spezifiziert werden, in welchem neben der Gleichsetzung der Ladungen auch die Intercepts über die Zeit invariant bleiben und somit die Gleichheitsrestriktionen auf skalarer Ebene der Invarianz erreicht werden.

4.3.3 Skalare Invarianz – DEMAT

Bedingung für die skalare Invarianz der Modellspezifikation ist es, dass Gleichheitsrestriktionen für Intercepts der Faktoren zugelassen werden können. Die Gleichheitsrestriktionen der Faktoren bleiben wie im Modell mit metrischer Invarianz erhalten, ebenso die Fixierung der Intercepts des 1. Indikators (Subtest Arithmetik) auf null, jedoch werden zusätzlich die Intercepts des Geometriebereichs fixiert, wie auch die Subskalen des Bereichs Sachrechnens. Im Anschluss werden die Schätzwerte für die latenten Variablen angefordert.

Die Frage der Fixierbarkeit der Schwierigkeiten der Geometrie- und Sachrechenaufgaben zeigt eine nicht zufriedenstellende Modellpassung. Der Wert des χ^2 weist 975.000 aus und 51 Freiheitsgrade, der CFI beträgt .81.

Die Schwierigkeiten der Aufgaben der Klassenstufe 3 und 4 weisen jedoch einen hohen Grad der Ähnlichkeitsbeziehung auf und lassen sich daher gut fixieren. Bei der Schätzung des Intercepts zeigt sich leider kein gutes Ergebnis. Daher werden lediglich die Sachrechenaufgaben fixiert. Für die Fixierung der Sachrechenaufgaben lässt sich nunmehr eine gute Modellpassung nachweisen: χ^2 beträgt 98.892 mit 46 Freiheitsgraden, sodass mit Einschränkungen eine recht akzeptable, starke faktorielle Invarianz auf skalarem Niveau erreicht werden kann. Der SRMR beträgt .03 und der RMSEA .02. Die quadrierte Faktorladung liegt mit .17 am niedrigsten im Geometriebereich des DEMAT 1 und mit einem Wert von .84 in der Arithmetikskala des DEMAT 2 am höchsten.

4.3.4 Invarianz Modellvergleich – DEMAT

Die Analyse der χ^2 -Quadrat-Differenztestung erfolgt auf der theoretischen Grundlage von Geiser (2010) für die deskriptive Statistik für Modellvergleiche. Die auch als Likelihood-Quotienten-Tests bezeichneten Prüfverfahren testen die Passung von ineinander geschachtelten Strukturgleichungsmodellen (Eid et al., 2013). Die Ergebnisse sind in der folgenden Tabelle dargestellt.

Tabelle 26

 χ^2 -Differenzentest DEMAT – Modellvergleich

	Modell unter konfiguraler Invarianzannahme	Modell unter metrischer Invarianzannahme mit 4 Residualkovarianzen	Modell unter partiell-metrischer Invarianzannahme mit 4 Residualkovarianzen	Modell unter skalarer Invarianzannahme mit 4 Residualkovarianzen
	vs.	vs.	vs.	vs.
	Modell unter konfiguraler Invarianzannahme mit 4 Residualkovarianzen	Modell unter konfiguraler Invarianzannahme mit 4 Residualkovarianzen	Modell unter metrischer Invarianzannahme mit 4 Residualkovarianzen	Modell unter partiell-metrischer Invarianzannahme mit 4 Residualkovarianzen
$\Delta\chi^2$	12.616	533.02	-533.02	22.30
Δdf	14	6	-5	1
$p (\chi^2\text{-Test})$.56	.00	keine Konvergenz	.00

$\Delta\chi^2$ = Differenz; Δdf = Differenz der Freiheitsgrade; p = Signifikanz.

Die χ^2 -Statistik führt zu Ablehnung der Nullhypothese, dass das Modell perfekt passen könnte. Die von Muthén und Muthén (1998-2007) konzipierte Analysesoftware erlaubt Analysen von hierarchisch geschachtelten Modellen. Es werden schrittweise Restriktionen zugelassen, wodurch das restriktivere Modell mehr Freiheitsgrade durch zusätzliche Beschränkungen im Vergleich zum weniger restriktiven Modell besitzt. Aufgrund dieser Einschränkungen passt das restriktivere Modell in der Regel weniger gut auf die Daten.

Nachfolgend sollen in der untenstehenden Tabelle die Passungswerte der einzelnen Modelle im Überblick vorgestellt werden.

Tabelle 27

Passung der einzelnen Modelle für DEMAT

	Modell unter konfiguraler Invarianzan- nahme	Modell unter konfiguraler Invarianz- annahme	Modell unter partiell-metrischer Invarianzannahme	Modell unter skalarer Inva- rianzannahme
		mit 4 Residualkovarianzen		
χ^2	63.99	76.60	609.62	76.60
df	30	44	50	45
p	.000	.002	.000	.002
CFI	.99	.99	.88	.99
$RMSEA$.03	.02	.08	.02

χ^2 = Diskrepanz; df = Freiheitsgrade; p = Signifikanz; CFI = Comparative Fit Index; $RMSEA$ = Root Mean Square Error of Approximation.

In folgender Tabelle findet sich die Übersicht über die Modellvergleiche der getesteten Modelle zur Erfassung der Messinvarianz mit vs. ohne Kontrolle der Mehrebenenstruktur, da die Daten geclustert in Schulen erhoben worden sind. Insgesamt kann ausgesagt werden, dass die Datenstruktur auch ohne Kontrolle der Mehrebenenstruktur gleich gut repräsentiert wird. Das favorisierte Modell 1b mit den vier zugelassenen Residualkovarianzen (ohne MEK) ist in der dritten Spalte zu finden.

Tabelle 28

Prüfung der Invarianz (jeweils ohne und mit Kontrolle der Mehrebenenstruktur)

Modell	1 a ohne MEK	1 a mit MEK	1 b ohne MEK	1 b mit MEK
	mit Residualkovarianzen		mit 4 Residualkovarianzen	
χ^2	63.99	56.66	76.60	70.35
<i>df</i>	30	30	44	44
<i>CFI</i>	.99	.99	.99	.99
<i>RMSEA</i>	.03	.02	.02	.02
<i>SRMR</i>	.02	.02	.02	.02

MEK = Mehrebenenkontrolle; χ^2 = Diskrepanz; *df* = Freiheitsgrade; *CFI* = Comparative Fit Index; *RMSEA* = Root Mean Square Error of Approximation (durchschnittliche Residuen); *SRMR* = standardisierte Residuen.

4.3.5 Invarianz – Zusammenfassung und Diskussion der Ergebnisse

Hinsichtlich der Messinvarianz lässt sich festhalten, dass die Messung kognitiver Fähigkeiten und die Messung mathematischer Schulleistungen über die Zeit invariant sind, allerdings, für die Messung der Testverfahrenreihe des DEMAT nur partiell. Die nur partielle Invarianz der Mathematikleistungen über die Zeit ist durch unterschiedliche Testformen inhaltlich erklärbar.

Damit kann die Hypothese 2 für die Erwartung der Invarianz für das Testverfahren des KFT bestätigt werden. Die Stabilitätskoeffizienten sehr gut sind. Die Faktorladungen des KFT 1 bis 4 mit den quantitativen Subskalen, den verbalen und nonverbalen Subskalen zeigen auf latenter Ebene durchweg hohe bis sehr hohe Korrelationen, was als sehr zufriedenstellend interpretiert werden kann. Damit ist für den KFT die Invarianz gut abgesichert. Mit diesem Ergebnis kann die erwartete Hypothese, dass die Invarianz für die KFT-Reihe über die vier Messzeitpunkte erfüllt ist, bestätigt werden. Als geeignetstes Modell erwies sich das Modell mit konfiguraler Invarianz. Zusammenfassend weisen die verwendeten Skalen zur Erfassung der kognitiven Fähigkeiten eine hohe Messgenauigkeit und gute Indi-

katorspezifität über die Zeit hinweg auf. Die Skalenparameter haben sich nur gering verändert, die stärkste Veränderung gab es zwischen dem zweiten und dritten Untersuchungsjahr. Messfehlerbereinigte interindividuelle Unterschiede hinsichtlich kognitiver Leistungen der untersuchten Schülerinnen und Schüler sind relativ stabil über die Messzeitpunkte.

Aus statistischer Sicht ist eine geringere Residualkovarianz auf latenter Ebene günstiger zu bewerten, jedoch sollten sich diese nicht als Messfehlerkorrelation finden. Inhaltlich aber gibt es in den überprüften Testverfahren gute Gründe für die Nichtkorrelation der Messfehler. Die Messverfahren wechseln zwischen dem zweiten und dritten Messzeitpunkt und daher ist eine Nichtkorrelation der Messfehler durchaus positiv zu beurteilen.

Die Hypothese der Erwartung von gegebener Invarianz der Testverfahren des DEMAT kann nur teilweise bestätigt werden. Die verwendeten Skalen zur Erfassung der mathematischen Fähigkeiten weisen eine hohe Messgenauigkeit und eine geringe Indikatorspezifität über die Zeit hinweg. Die Skalenparameter haben sich nur gering verändert, die stärkste Veränderung gab es zwischen dem zweiten und dritten Untersuchungsjahr. Messfehlerbereinigte interindividuelle Unterschiede hinsichtlich mathematischer Leistungen der untersuchten Schülerinnen und Schüler sind relativ stabil über die Messzeitpunkte. Die latenten State-Faktoren lassen eine Korrelationsstruktur erkennen, die auf autoregressive Effekte hinweisen. Zwar gab es einige Unterschiede in den Faktorindikatoren in den verschiedenen Klassenstufen. In der Konstruktion sind die Subtests einiger Skalen jedoch explizit als Anwärmaufgaben deklariert worden und besitzen daher eine geringere Trennschärfe. Damit sind die Zusammenhänge nicht sehr hoch, aber durchaus im moderaten Bereich und können die Zusammenhangsstruktur ausweisen. Damit ist die Messinvarianz des DEMAT über einen Vierjahreszeitraum lediglich partiell gegeben.

Es hat bislang nur wenige Studien gegeben, die interindividuelle Unterschiede zwischen kognitiven Fähigkeiten und mathematischen Schulleistungen betrachten. Dabei ist es wichtig zu beachten, dass die Messinstrumente, die in Längsschnittstudien eingesetzt werden, präzise dasselbe Konstrukt erfassen können, ohne dass Messfehler und Testverfahrenswechsel als tatsächliche Veränderung in die Ergebnisse einfließen und diese unter Umständen entwicklungsbedingten Veränderungen oder den Effekten von Traitmaßnahmen zugeschrieben werden. Die Konstruktkonsistenz der im Grundschulalter eingesetzten Testverfahren bestimmt wesentlich die Höhe der Messfehler und sollte bereits vor dem Einsatz berücksichtigt werden, sofern längsschnittliche Fragestellungen beantwortet werden. Auch bei der Bewertung vormaliger Längsschnittdaten zu diesem Thema ist die Berücksichtigung

der Messfehler qualitätsbestimmend. Um zuverlässige Aussagen aus Längsschnittdaten abzuleiten, sollten Längsschnittstudien situationsspezifische bzw. messfehlerbedingte Einflüsse berücksichtigen, deren unerwünschte Kumulationen bei Unterlassung der Kontrolle über einen längeren Untersuchungszeitraum zu verzerrten Schlussfolgerungen führen können.

4.4 Ergebnisse 3 – Autoregressive Effekte

Entsprechend des in Kapitel 3 beschriebenen Analysedesigns werden nun die Ergebnisse der autoregressiven Modellspezifikationen berichtet. Dabei erfolgt die Beschreibung der Ergebnisse der Spezifikationen der Modelle schrittweise, zunächst für das Modell erster Ordnung, anschließend für das Modell der zweiten Ordnung und nachfolgend für das Modell der dritten Ordnung.

4.4.1 Autoregressive Effekte – 1. Ordnung

Zunächst werden die Ergebnisse aufgrund des Modells 1. Ordnung dargestellt, die jeweils konstruktsspezifische Regressionen auf den jeweils folgenden Messzeitpunkt ermitteln. In der nachfolgenden Abbildung werden die Regressionen für die Konstrukte separat veranschaulicht.

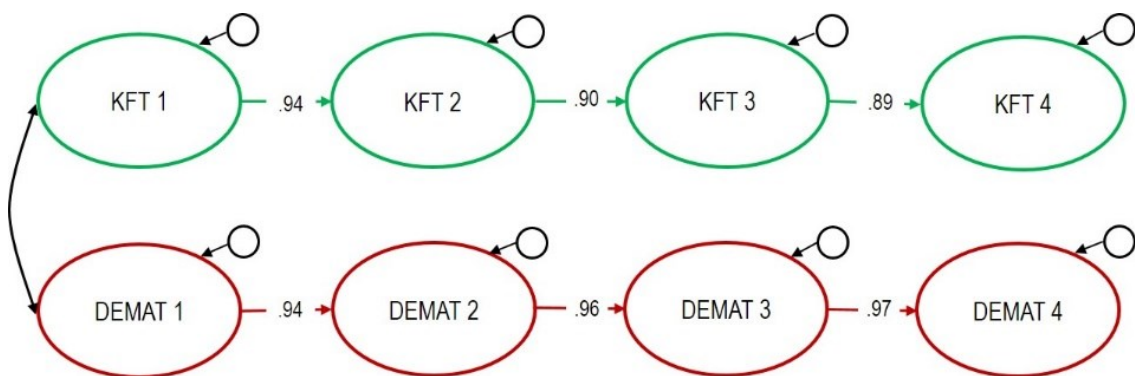


Abbildung 38 Autoregressives Modell 1. Ordnung.

Die autoregressiven Pfade des KFT 2 auf den KFT 1 der 1. Ordnung weisen einen Wert von .94 aus. Die Ergebnisse des KFT 3 können zu .90 auf den KFT 2 zurückgeführt werden und auch die Regression des KFT 4 auf den KFT 3 ist mit .89 sehr hoch.

In der Betrachtung der autoregressiven Pfade der mathematischen Fähigkeiten kann der DEMAT 2 auf den DEMAT 1 mit .94 zurückgeführt werden. Die Vorhersage des

DEMAT 2 auf den DEMAT 3 gelingt mit .96 und die Regression des DEMAT 4 auf den DEMAT 3 mit .97. Mit diesen Analyseergebnissen kann festgestellt werden, dass die inter-individuelle Stabilität der jeweiligen Leistungen vom 1. Messzeitpunkt bis zum 4. Messzeitpunkt und damit über die vier Grundschuljahre hoch ist.

4.4.2 Autoregressive Effekte – 2. Ordnung

Im Anschluss soll analysiert werden, inwieweit sich Ergebnisse aus den früheren Messungen über zwei Messzeitpunkte hinweg vorhersagen lassen. Explizit interessiert, inwieweit die KFT 3 Ergebnisse erstens noch aus den KFT 1 Testergebnissen vorhersagen lassen und zweitens, inwieweit sich KFT 4 Ergebnisse aus den Messungen des KFT 2 vorhersagen lassen. Analog erfolgt die Spezifikation der Konstrukte mathematischer Schulleistungen über zwei Messzeitpunkte hinweg. Die folgende Tabelle weist die Ergebnisse aus.

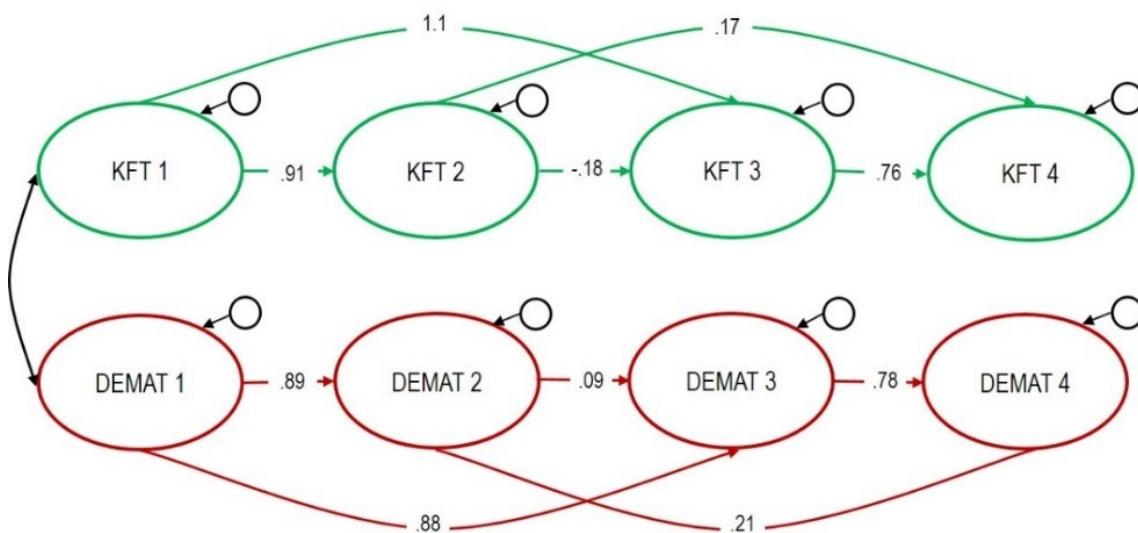


Abbildung 39 Autoregressives Modell 2. Ordnung.

Die autoregressiven Pfade des KFT 2 auf den KFT 1 der 1. Ordnung weisen einen Wert von .91 aus. Die Regression des KFT 4 auf den KFT 3 ist mit .76 recht gut. Die Pfade des KFT 3 können zu -.18 auf den KFT 2 zurückgeführt werden und die autoregressiven Pfade der zweiten Ordnung stellen sich innerhalb des KFT wie folgt dar: KFT 3 auf KFT 1 mit 1.1 und KFT 4 auf KFT 2 mit .17, was auf eine nicht optimale Modellpassung hinweist.

In der Betrachtung der autoregressiven Pfade der mathematischen Fähigkeiten kann der DEMAT 2 auf den DEMAT 1 mit .89 zurückgeführt werden. Die Vorhersage des

DEMAT 2 auf den DEMAT 3 gelingt mit .09 und die Regression des DEMAT 4 auf den DEMAT 3 mit .78. Die Pfade der Effekte innerhalb des Konstrukts mathematische Fähigkeiten zeigen auf zweiter Ordnungsebene folgende Ergebnisse: DEMAT 3 auf DEMAT 1 kann einen akzeptablen Einfluss von .88 nachweisen. Die Vorhersage des DEMAT 4 aus dem DEMAT 2 gelingt mit .21.

4.4.3 Autoregressive Effekte – 3. Ordnung

Mittels des darauffolgenden Modells soll nun geprüft werden, inwiefern sich zusätzlich die KFT 4 Ergebnisse aus den KFT 1 Ergebnissen vorhersagen lassen und weiter, inwieweit die DEMAT 4 Ergebnisse aus den DEMAT 1 Messungen vorhergesagt werden können.

Die autoregressiven Pfade des KFT 2 auf den KFT 1 der 1. Ordnung weisen einen Wert von .90 aus. Die Ergebnisse des KFT 3 können mit nunmehr einem Wert von .05 kaum auf den KFT 2 zurückgeführt werden, resultieren die Ergebnisse jedoch vornehmlich aus den Werten des KFT 1. Ein Ergebnis, das sich bereits im Modell zweiter Ordnung andeutete, aber dennoch in dieser Spezifikation eine deutlich bessere Passung zeigt. Auch die Regression des KFT 4 auf den KFT 3 ist mit .13 recht gut, wie der folgenden Abbildung zu entnehmen ist.

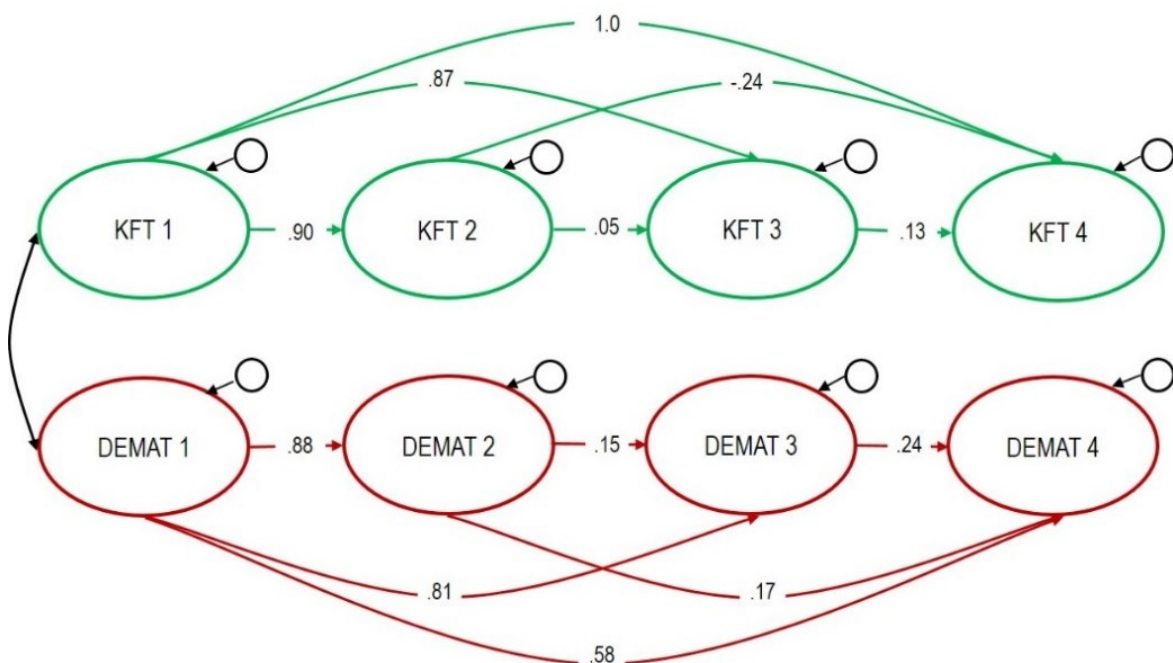


Abbildung 40 Autoregressives Modell der 3. Ordnung.

Die autoregressiven Pfade der zweiten Ordnung stellen sich innerhalb des KFT wie folgt dar. Die Regressionen des KFT 4 auf den KFT 2 sind mit .24 moderat und die des KFT 3 auf den KFT 1 ist mit .87 sehr hoch.

Die Untersuchung der 3. Ordnung, der Effekte über zwei zurückliegende Messungen, zeigen innerhalb der kognitiven Fähigkeiten in der Darstellung der Regressionen des KFT 4 auf den KFT 1 einen Wert über 1 und sind nicht interpretierbar.

In der Betrachtung der autoregressiven Pfade der mathematischen Schulleistungen kann der DEMAT 2 auf den DEMAT 1 mit .88 zurückgeführt werden. Die Vorhersage des DEMAT 2 auf den DEMAT 3 gelingt mit .15 und die Regression des DEMAT 4 auf den DEMAT 3 mit .24.

Die Pfade der Effekte innerhalb des Konstrukts mathematische Leistungen zeigen auf zweiter Ordnungsebene folgende Ergebnisse: DEMAT 3 auf DEMAT 1 kann einen akzeptablen Einfluss von .81 nachweisen. Die Vorhersage des DEMAT 4 aus dem DEMAT 2 gelingt mit .17.

Die konstruktinternen Regressionen der dritten Ordnung der DEMAT-Testverfahren zeigen den Einfluss des DEMAT 1 auf den DEMAT 4 in Höhe von .58. Für die Mathematikleistungen sind ebenfalls die Regression von DEMAT 2 auf DEMAT 1 mit .88 hoch und in der Betrachtung der 2. Ordnung wirken DEMAT-Ergebnisse des 1. Messzeitpunktes auf Mathematikleistungen des 3. Messzeitpunktes mit .81.

Zu erkennen ist, dass es zwischen dem DEMAT 2 und DEMAT 3 sowie zwischen KFT 2 und KFT 3 nur geringe Wirkbeziehungen gibt. Bereits bei der Prüfung der Invarianz ergab die Prüfung lediglich partielle Invarianz. Daher können diese Ergebnisse inhaltlich mit dem Wechsel der Testformen und durch die Orientierung am jeweiligen Curriculum des Jahrgangs und nicht an der Orientierung der Testautoren an der Konstruktkonsistenz über die vier Grundschuljahre, erklärt werden (Krajewski, 2003). Dennoch weisen die Spezifikationen der autoregressiven Pfade der mathematischen Leistungen auf eine relativ hohe Stabilität der Merkmalsausprägungen von Messungen aus Vorjahren auf Ergebnisse in Folgejahren hin.

Zusammenfassend kann festgehalten werden, dass die autoregressiven Pfade der 1., 2. und 3. Ordnung auf eine relativ hohe Stabilität der Merkmalsausprägungen von Messungen aus Vorjahren auf Ergebnisse in Folgejahren hinweisen.

4.4.3.1 Autoregressive Effekte – Passungswerte

Die untenstehende Tabelle zeigt die Passungswerte der untersuchten autoregressiven Effekte. Das autoregressive Modell der dritten Ordnung weist mit einem χ^2 von 1043.82 und 216 sowie einem CFI in Höhe von .95 im Vergleich die deutlich bessere Passung auf.

Tabelle 29

Passungswerte für autoregressive Modelle – Stabilität

	<i>ARM 1. Ordnung</i>	<i>ARM 2. Ordnung</i>	<i>ARM 3. Ordnung</i>
χ^2	1348.30	1121.12	1043.82
<i>df</i>	222	218	216
<i>CFI</i>	.93	.94	.95
<i>RMSEA</i>	.06	.05	.05
$\Delta\chi^2$		227.18	77.30
Δdf		4	2

ARM = Autoregressives Modell; χ^2 = Diskrepanz; $\Delta\chi^2$ = Differenz Diskrepanzfunktion; *df* = Freiheitsgrade; Δdf = Differenz Freiheitsgrade; *CFI* = Comparative Fit Index; *RMSEA* = Root Mean Square Error of Approximation.

4.4.3.2 Autoregressive Effekte – Kontrolle der Mehrebenenstruktur

Da die Stichprobe der Kinder in Klassen und Schulen geclustert vorlag, wurde die Mehrebenenstruktur kontrolliert und die Modelle mit denen ohne die Kontrolle der mehrebenenstruktur verglichen. Wie die untenstehende Tabelle zeigt, zeigt der Vergleich keine bedeutenden Unterschiede.

Tabelle 30

Passungswerte angewandter Modelle zur Prüfung autoregressiver Effekte (Modelle 1 bis 3) jeweils ohne und mit Kontrolle der Mehrebenenstruktur

Modell	1 ohne MEK	1 mit MEK	2 ohne MEK	2 mit MEK	3 ohne MEK	3 mit MEK
χ^2	1348.30	1219.14	1121.12	1031.62	1043.82	940.74
<i>df</i>	222	222	218	218	216	216
<i>CFI</i>	.93	.93	.94	.94	.95	.95
<i>RMSEA</i>	.06	.05	.05	.05	.05	.04
<i>SRMR</i>	.68	.07	.04	.04	.04	.04

MEK = Mehrebenenkontrolle; χ^2 = Diskrepanz; *df* = Freiheitsgrade; *CFI* = Comparative Fit Index; *RMSEA* = Root Mean Square Error of Approximation (durchschnittliche Residuen); *SRMR* = standardisierte Residuen.

Im folgenden Kapitel sollen die wichtigsten Ergebnisse der Prüfung autoregressiver Effekte zusammengefasst werden.

4.4.4 Autoregressive Effekte – Zusammenfassung und Diskussion der Ergebnisse

In der Frage der Stabilität kognitiver Fähigkeiten und mathematischer Schulleistungen kann zusammenfassend auf der Ebene der Konstrukte davon ausgegangen werden, dass über den betrachteten Vierjahreszeitraum sowohl eine hohe Stabilität der kognitiven Fähigkeiten als auch eine hohe Stabilität mathematischer Leistungen im Grundschulalter festzustellen ist. Die Aussagen für das Grundschulalter können als präzise gelten, da vier Messzeitpunkte aus unabhängigen Stichproben herangezogen werden konnten. Nur bei einzelnen Kindern zeigen sich Verschiebungen in den Rangreihen.

Die autoregressiven Pfade des kognitiven Fähigkeitstests 1. Ordnung (KFT 2 auf KFT 1 sowie KFT 3 auf KFT 2 als auch KFT 4 auf KFT 3) der 2. Ordnung (KFT 3 auf den KFT 1 sowie des KFT 4 auf KFT 2), als auch der 3. Ordnung (KFT 4 auf KFT 1) lassen gute bis akzeptable Rückschlüsse auf die Stabilität der Merkmalsmessung zu. Die Spezifikation der autoregressiven Pfade der mathematischen Fähigkeiten weist ebenso eine recht

gute Vorhersagekraft früherer Ergebnisse auf Messungen in Folgejahren auf, wobei für Geometrie die geringsten Zusammenhänge nachgewiesen wurden.

4.5 Ergebnisse 4 – Kreuzverzögerte Effekte

4.5.1 Kreuzverzögerte Effekte – Einflüsse kognitiver Fähigkeiten auf spätere mathematische Schulleistungen

Zunächst sollen die Einflüsse kognitiver Fähigkeiten aus Vorjahren auf spätere mathematische Schulleistungen ermittelt werden. Dafür wird zunächst das autoregressive Modell der dritten Ordnung als Basismodell herangezogen. Das nunmehr spezifizierte Modell, das kreuzverzögerte Effekte darstellen kann, erfolgt auf der Grundlage der Zulassung der Regressionspfade der kognitiven Fähigkeiten früherer Zeitpunkte auf die jeweils folgenden Messzeitpunkte der mathematischen Schulleistungen. Das sogenannte Crossed-Modell der 1. Ordnung in der folgenden Abbildung beschreibt zunächst die Einflüsse kognitiver Fähigkeiten auf die mathematischen Leistungen.

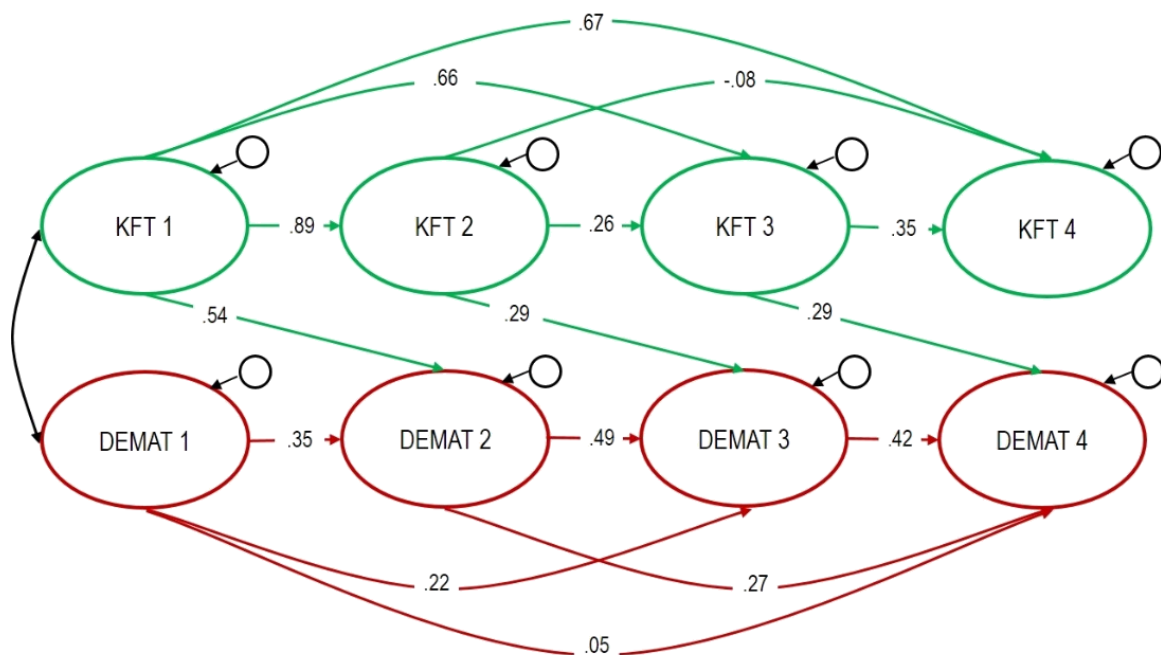


Abbildung 41 Crossed-Lagged-Effekte 1. Ordnung (Modell 4).

In dem Crossed-Lagged-Modell der 1. Ordnung (Modell 4), in welchem sowohl autoregressive Effekte zwischen den Messzeitpunkten des gleichen Konstrukts betrachtet wurden, als auch kreuzverzögerte Effekte der kognitiven Fähigkeitstests auf den jeweils folgenden

Messzeitpunkt der mathematischen Schulleistungen im Vordergrund stehen, zeigen sich folgende Ergebnisse:

Die autoregressiven Ergebnisse kognitiver Fähigkeiten zeigen auf der ersten Ordnungsebene, dass die Ergebnisse des KFT 2 mit .89 auf die kognitiven Fähigkeiten des ersten Messzeitpunktes zurückgeführt werden können. Im Vergleich zu der vorherigen Spezifikation in einem autoregressiven Modell ohne kreuzverzögernde Einflüsse zeigt sich eine deutliche Verbesserung der Modellpassung. Im Vergleich zu dem autoregressiven Modell (3) verbessert sich die Modellpassung signifikant, wenn die Pfade des jeweils früheren Prädiktors kognitiver Fähigkeiten (KFT) auf die mathematischen Schulleistungen (DEMAT) Leistung hinzugenommen werden.

Kognitive Fähigkeiten des dritten Messzeitpunktes lassen sich mit .26 auf kognitive Fähigkeiten des zweiten Messzeitpunktes zurückführen. Dieses relativ mäßige Ergebnis kann inhaltlich mit dem Wechsel der Messverfahren der Autoren des kognitiven Fähigkeitstests erklärt werden. Die Regression der kognitiven Fähigkeiten des dritten Messzeitpunktes auf die kognitiven Fähigkeiten des ersten Messzeitpunktes zeigen jedoch durchaus sehr ansprechende Ergebnisse mit .66.

Der kognitive Fähigkeitstest für die vierte Klassenstufe lässt sich im Modell dritter Ordnung mit .35 aus dem kognitiven Fähigkeitstest für die dritte Klassenstufe vorhersagen. Hingegen lässt sich der Einfluss des kognitiven Fähigkeitstests für die dritte Klassenstufe auf den kognitiven Fähigkeitstest des vierten Messzeitpunktes in diesem Modell nicht nachweisen, wobei die Vorhersage des kognitiven Fähigkeitstests des vierten Messzeitpunktes aus den Ergebnissen des kognitiven Fähigkeitstests des ersten Messzeitpunktes mit .67 ausgesprochen hoch ist.

Im weiteren Verlauf werden die autoregressiven Ergebnisse der mathematischen Schulleistungen vorgestellt. Der DEMAT des zweiten Messzeitpunktes lässt sich im Modell 1. Ordnung mit .35 auf die Ergebnisse des DEMAT des ersten Messzeitpunktes zurückführen und der DEMAT des dritten Messzeitpunktes mit .49 auf den DEMAT des zweiten Messzeitpunktes. Der DEMAT 4 kann zu .42 aus den DEMAT 3-Ergebnissen vorhergesagt werden.

Auf der 2. Ordnungsebene kann der DEMAT 3 mit .22 aus dem DEMAT 1 erklärt werden und der DEMAT 4 mit .27 aus den Ergebnissen des DEMAT 2.

Auf der dritten Ordnungsebene lassen sich die Ergebnisse des DEMAT 4 nur gering auf die Ergebnisse des DEMAT 1 zurückführen, was sich inhaltlich jedoch gut erklären

lässt, da die interne Konsistenz zwischen DEMAT 1 und DEMAT 4 dem Wechsel der Testverfahren und damit geringen, aber vorhandenen Konzeptionsabweichungen geschuldet ist.

Die Hypothese 4 kann folglich bestätigt werden, da festgestellt werden konnte, dass sich mathematische Leistungen auf kognitive Fähigkeiten zurückführen lassen.

Die Pfade zeigen eine vorhandene Beeinflussung von kognitiven Fähigkeiten aus Vorjahren auf Mathematikleistungen zu späteren Zeitpunkten. Die kognitiven Fähigkeiten des 1. Messzeitpunktes haben mit .54 recht hohen Einfluss auf die mathematischen Schulleistungen des 2. Messzeitpunktes. Der Einfluss von KFT 2 auf DEMAT 2 ist mit .29 moderat, ebenso wie DEMAT 4-Ergebnisse mit einem Regressionsgewicht von .29 auf kognitive Fähigkeiten der 3. Klassenstufe zurückzuführen sind.

4.5.2 Kreuzverzögerte Effekte – Hinzunahme der Einflüsse mathematischer Schulleistungen auf kognitive Fähigkeiten

Um nun die kreuzverzögerten Effekte zu ermitteln, wird das beschriebene Modell erweitert, indem nun auch die Regressionspfade der Messungen des DEMAT früherer Messzeitpunkte auf die kognitiven Fähigkeiten des jeweils folgenden Messzeitpunktes zusätzlich hinzugenommen werden. Die kreuzverzögerten Effekte in dem Crossed-Modell der 2. Ordnung zeigen folgende in der Abbildung dargestellten Ergebnisse.

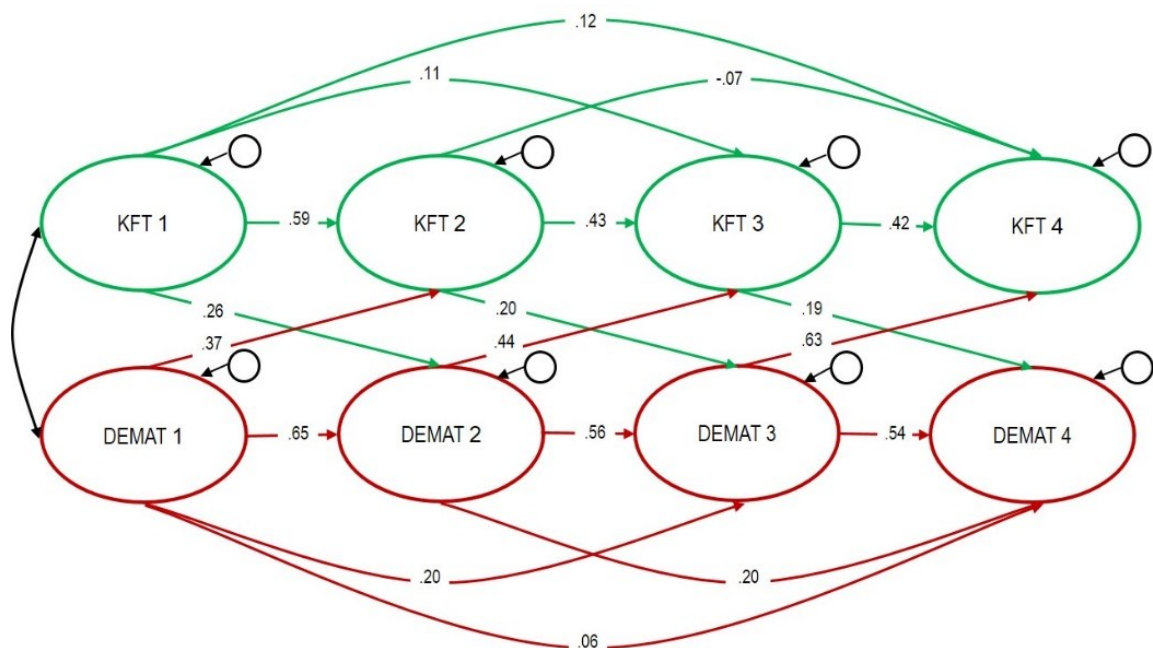


Abbildung 42 Crossed-Lagged-Effekte 2. Ordnung (Modell 5).

Zunächst werden die autoregressiven Ergebnisse des KFT unter Zulassung dieser Erweiterung berichtet. Die Ergebnisse des KFT 2 können mit .59 auf den KFT 1 zurückgeführt werden. Der KFT 3 lässt sich mit .43 auf den KFT 2 zurückführen, was im Vergleich zum vorherigen Modell eine deutliche Verbesserung darstellt. Der KFT 4 lässt sich im Modell dritter Ordnung mit .42 aus dem KFT 3 vorhersagen und damit zu einem höheren Anteil als im Crossed-Lagged-Modell 4.

Die Regression des KFT 3 auf den KFT 1 zeigt mit .11 jedoch nur einen geringen Einfluss, was sich jedoch wiederum inhaltlich begründet mit dem Wechsel des Testverfahrens erklären lässt.

Die Leistungen, die im KFT 4 erbracht wurden, lassen sich in diesem Modell kaum auf Leistungen des zum KFT 2 zurückführen, wobei die Vorhersage des KFT 4 aus den Ergebnissen des KFT 1 mit .12 gelingt.

Im Anschluss erfolgt die Darstellung der autoregressiven Ergebnisse des DEMAT. Der DEMAT 2 lässt sich im Modell 1. Ordnung mit .65 auf die Ergebnisse des DEMAT 1 zurückführen, und damit im Vergleich zum Modell 4 eine deutliche Verbesserung darstellt. Der DEMAT 3 kann mit .56 auf den DEMAT 2 zurückgeführt werden, was ebenfalls eine deutliche Verbesserung zum Vormodell darstellt. Der DEMAT 4 kann zu .54 aus den DEMAT 3 Ergebnissen vorhergesagt werden. Auch hier zeigt sich eine deutliche Verbesserung der Datenpassung.

Auf der 2. Ordnungsebene kann der DEMAT 3 mit .20 aus dem DEMAT 1 erklärt werden, was nur eine geringe Veränderung im Vergleich zu Vormodell darstellt. Der DEMAT 4 kann mit .20 aus den Ergebnissen des DEMAT 2 vorhergesagt werden.

Auf der dritten Ordnungsebene lassen sich die Ergebnisse des DEMAT 4 nur mit .06 auf die Ergebnisse des DEMAT 1 zurückführen, was sich inhaltlich jedoch gut erklären lässt, da die interne Konsistenz zwischen Testverfahren DEMAT 1 und DEMAT 4 durch den bereits erwähnten Wechsel der Testadaptionen und damit geringen, aber vorhandenen Konzeptionsabweichungen unterworfen ist.

Die Ergebnisse der Crossed-Lagged-Effekte der ersten Ordnung unter Hinzuziehung der gegenseitigen Regressionen (DEMAT vorhergesagt durch KFT, KFT vorhergesagt durch DEMAT) zeigen folgendes Bild.

Der KFT 2 lässt sich zu .37 aus dem DEMAT 1 erklären, während sich der DEMAT 2 zu .26 aus dem KFT 1 erklären lässt. Der KFT 3 lässt sich mit einem Wert von .44 aus dem DEMAT 2 erklären, während der DEMAT 3 sich zu .20 aus dem KFT 2 vorhersagen

lässt. Der KFT 4 regrediert zu .63 auf dem DEMAT 3 während der DEMAT 4 sich zu .19 aus dem KFT 3 vorhersagen lässt.

Auffallend ist, dass sich die Fähigkeiten des DEMAT eher als Prädiktor der kognitiven Fähigkeiten auswirken, als sich die mathematischen Fähigkeiten aus den kognitiven Fähigkeiten erklären lassen.

Die Untersuchung der Vorhersagekraft der kognitiven Fähigkeiten durch die Messung der mathematischen Schulleistung konnte im 5. Modell als Crossed-Lagged-Modell weiter spezifiziert werden. Dabei beschränkt sich die Konzeption der Regressionen nur auf die direkt vorangegangene Messung. Es zeigt sich im Ergebnis ein χ^2 von 659.59 und 210 Freiheitsgraden bei einem sehr guten RMSEA von .05 und einem SRMR von .03. Die Beurteilung der Modellgüte kann mit einem CFI von .97 als sehr gut bezeichnet werden.

4.5.3 Kreuzverzögerte Effekte – unter Hinzuziehung der komplexen Beziehungsmuster

Das im nachfolgenden Schritt spezifizierte Modell (folgende Abbildung) ist dann das angestrebte Modell (Modell 6), das beidseitige Einflüsse aus vorherigen Messungen auf Folgeleistungen des anderen Konstrukts ausgeben soll. Dabei erfolgt die Berücksichtigung der regressiven Beziehungen unter gleichzeitiger Berücksichtigung auch der autoregressiven Beziehungen der dritten Ordnung, wie in der dargestellten Grafik gezeigt wird.

Hier interessiert nicht nur, ob die kognitiven Leistungen die kognitiven Leistungen späterer Zeitpunkte vorhersagen können, sondern inwieweit kognitive Fähigkeiten auf spätere mathematische Schulleistungen schließen lassen. Zusätzlich werden die Residualkovarianzen ermittelt und deren Korrelationen zum jeweils anderen Konstrukt ermittelt. Sie weisen konstante Muster auf, wie in der untenstehenden Abbildung ersichtlich.

Die vorher spezifizierten Regressionen des KFT zeigen weiterhin Einflüsse zum 1., 2. und 3. Folgejahr, analog auch bei den mathematischen Schulleistungen. Die Ergebnisse sind im abgebildeten Modell dargestellt.

Die roten Pfeile in der Mitte, ausgehend vom DEMAT 1 auf den KFT 2 und dann analog vom DEMAT 2 auf den KFT 3 sowie dann vom DEMAT 3 auf die kognitiven Leistungen des 4. Messzeitpunktes, veranschaulichen die kreuzverzögerte regressive Beziehung mathematischer Schulleistungen auf kognitive Fähigkeiten. Die grünen Pfade zeigen die Regressionen in die umgekehrte Richtung von kognitiven Fähigkeiten auf mathematische Schulleistungen auf die nachfolgenden Zeitpunkte.

Die Ergebnisse der autoregressiven Einflüsse können weiterhin bestätigt werden. Sie zeigen in der Ermittlung der 1. Ordnung, dass die Ergebnisse des KFT 2 mit .60 auf den KFT 1 zurückgeführt werden können, was im Vergleich zu den vorherigen Modellen eine weitere deutliche Verbesserung darstellt. Der KFT 3 lässt sich mit .40 auf den KFT 2 zurückführen. Der KFT 4 lässt sich mit .21 aus dem KFT 3 vorhersagen und gelingt damit etwas schlechter als im Crossed-Lagged-Modell 5. In der Betrachtung der 2. Ordnungsebene weist die Regression des KFT 3 auf den KFT 1 mit .17 einen immer noch geringen Einfluss auf, jedoch eine bessere Vorhersagekraft als im Vormodell.

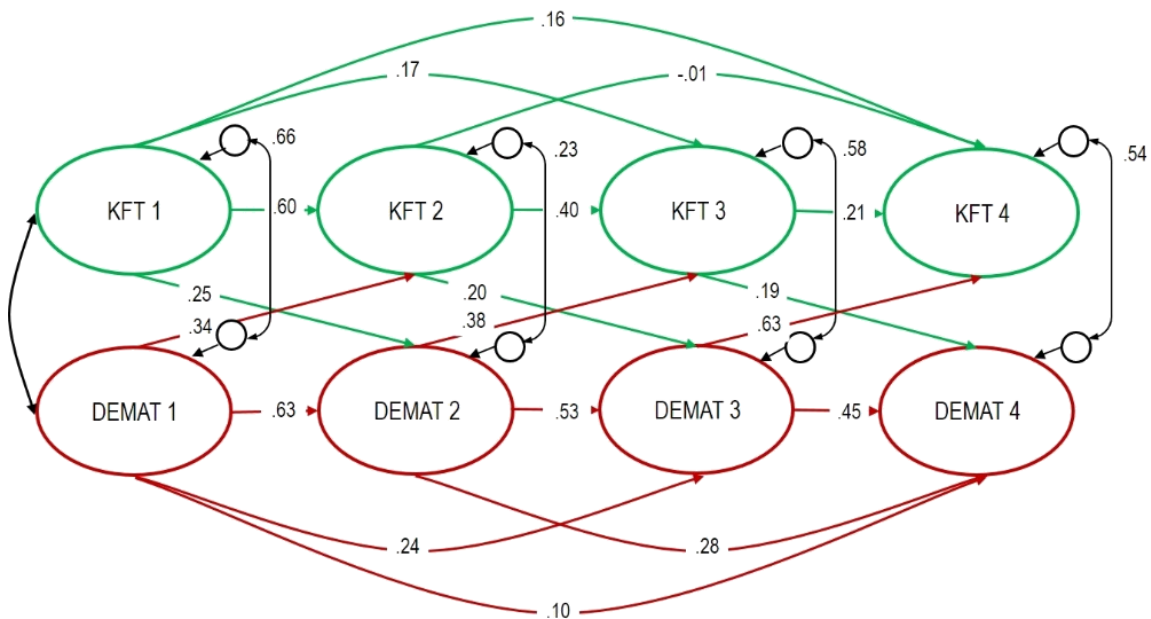


Abbildung 43 Crossed-Lagged-Effekte 3. Ordnung (Modell 6).

Die Vorhersagekraft des KFT 2 auf den KFT 4 gelingt weiterhin nicht. In der Betrachtung der 3. Ordnungsebene zeigt sich die Vorhersagemöglichkeit des KFT 4 aus den Ergebnissen des KFT 1 mit einem etwas höheren, aber dennoch geringen Wert von .16.

Die autoregressiven Ergebnisse des DEMAT-Testverfahrens über die Zeit zeigen, dass sich die DEMAT 2-Ergebnisse auf erster Ordnungsebene mit .63 auf die Ergebnisse des DEMAT 1 zurückführen lassen, was im Vergleich zum Modell 5 ein gleichbleibendes Ergebnis darstellt. Der DEMAT 3 kann zu .53 auf den DEMAT 2 zurückgeführt werden, was ebenfalls ein fast analoges Ergebnis zu Modell 5 darstellt. Der DEMAT 4 kann zu .45 aus den DEMAT 3 Ergebnissen vorhergesagt werden. Auch hier zeigt sich eine ähnliche Beziehung im Vergleich zu Modell 5.

Auf 2. Ordnungsebene kann der DEMAT 3 mit .24 aus dem DEMAT 1 erklärt werden, was wiederum eine geringe Verbesserung im Vergleich zu Vormodell darstellt. Der DEMAT 4 kann mit .28 aus den Ergebnissen des DEMAT 2 vorhergesagt werden. Auch dieses Ergebnis ist eine Verbesserung im Vergleich zum Modell 5.

Auf der dritten Ordnungsebene lassen sich die Ergebnisse des DEMAT 4 nunmehr mit .10 auf die Ergebnisse des DEMAT 1 zurückführen, was eine leichte Verbesserung trotz des bereits erwähnten Wechsels des Testverfahrens zwischen den Jahrgangsstufen darstellt und die geringe Beziehung erklärt.

Die Ergebnisse der Crossed-Lagged-Effekte der ersten Ordnung unter Hinzuziehung der gegenseitigen Regressionen (DEMAT vorhergesagt durch KFT, KFT vorhergesagt durch DEMAT) zeigen folgendes Bild. Der KFT 2 lässt sich zu .34 aus dem DEMAT 1 erklären, während sich der DEMAT 2 zu .25 aus dem KFT 1 erklären lässt. Der KFT 3 lässt sich mit einem Wert von .38 aus dem DEMAT 2 erklären, während der DEMAT 3 sich zu .20 aus dem KFT 2 vorhersagen lässt. Der KFT 4 regrediert zu .63 auf den DEMAT 3 während der DEMAT 4 sich zu .19 aus dem KFT 3 vorhersagen lässt.

Auffallend ist weiterhin, dass die Fähigkeiten des DEMAT eher als Prädiktor der kognitiven Fähigkeiten wirken, als dass sich die mathematischen Fähigkeiten aus den kognitiven Fähigkeiten erklären lassen.

Bei der Betrachtung der kreuzverzögerten Effekte des KFT auf den DEMAT zeigt sich die zunehmende Vorhersagekraft über die Jahre. Der KFT 2 lässt sich mit .34 auf den DEMAT 1 zurückführen, während sich der DEMAT 2 mit .38 auf den KFT 3 zurückführen lässt und der KFT 4 mit einem Wert von .63 auf den DEMAT 3 zurückgeführt werden kann.

In der kreuzverzögerten Betrachtung des DEMAT kann der DEMAT 2 auf den KFT 1 mit .25 zurückgeführt werden, der DEMAT 3 mit .20 auf den KFT 2 und der DEMAT 4 auf den KFT 3 mit .20.

Das Modell hat sich weiterhin verbessert bei einem χ^2 von 593.96 und 208 Freiheitsgraden, einem optimalen RMSEA von .03 und einem SRMR von .03. Der CFI ist mit einem Wert von .98 als sehr gut zu beurteilen. Damit zeigt sich eine wesentlich bessere Datenanpassung.

Die quadrierte Faktorladung der latenten Variablen fällt bei der Betrachtung des KFT (zwischen .74 und .86) höher aus als in den vorherigen Modellspezifikationen. Damit zeigen die Reliabilitätswerte eine hohe bis sehr hohe Ausprägung und zeugen von einer hohen zeitlichen Stabilität der Merkmale.

Auch für den DEMAT zeigen sich Reliabilitätswerte zwischen .67 und .91. Damit kann dieses latente Modell deutlich präziser die Varianz des zweiten, dritten und vierten Messzeitpunktes vorhersagen.

Die mathematischen Schulleistungen zum 1. Messzeitpunkt wirken auf kognitive Fähigkeiten der 2. Klassenstufe mit .34 und damit höher als KFT 1 auf den DEMAT 2 mit .25. Der DEMAT 2 wirkt auf KFT 3 mit .38 ebenso höher als kognitive Leistungen des 2. Messzeitpunktes auf mathematische Schulleistungen des 3. Messzeitpunktes. Bei der Betrachtung der Beziehung zwischen mathematischen Schulleistungen der 3. Klassenstufe ist der Einfluss auf kognitive Leistungen der 4. Klassenstufe mit .63 beachtlich. Die gegenteilige Wirkung von KFT-Leistungen auf mathematische Schulleistungen fällt mit .19 im Vergleich recht gering aus.

Unter Hinzunahme dieser Pfade zeigt dieses Modell einen deutlichen Zugewinn an fit und deutlicher Reduktion der χ^2 -Werte. Die Passung ist deutlich besser und es werden nur 5 Freiheitsgrade verloren.

Die kreuzverzögerten Pfade zeigen eine gegenseitige Beeinflussung von kognitiven Fähigkeiten und Mathematikleistungen aus Vorjahren auf Ergebnisse in Folgejahren. Damit wird an dieser Stelle ein bedeutsamer Einfluss kreuzverzögerter Effekte deutlich, insbesondere in Richtung von mathematischen Schulleistungen auf kognitive Fähigkeiten, wo der Einfluss im Laufe der Grundschulzeit sogar noch zunimmt.

Zu erkennen ist auch, dass es zwischen dem DEMAT 2 und 3 sowie zwischen KFT 2 und 3 nur geringe Wirkbeziehungen gibt. Bereits bei der Prüfung der Invarianz ergab die Prüfung nur partielle Invarianz, das heißt, diese Ergebnisse können inhaltlich mit der Änderung des Konstruktverständnisses durch den Wechsel der Testformen erklärt werden. Es lässt sich erkennen, dass die Stabilität des Merkmals über die vier Jahre hoch ist und die Merkmalskriterien sich innerhalb der Messungen nicht wesentlich verändern.

Bei der Betrachtung der spezifischen Fähigkeiten, können die Ergebnisse inhaltlich begründet werden, da der DEMAT rein mathematische Leistungen erfasst, die in die Ebenen Geometrie, Arithmetik und Sachrechnen gegliedert sind. Der KFT bildet insgesamt ein breites Spektrum der kognitiven Fähigkeiten ab und kann bereichsspezifische Fähigkeiten innerhalb des Konstrukts der Intelligenz erfassen, wobei beispielsweise in dem Subtest verbale Fähigkeiten erfasst werden, die sich innerhalb des DEMAT nur gering auswirken dürften. In der Subskala nonverbale Fähigkeiten werden wahrnehmungsgebundene Fähigkeiten abgebildet, die dem Konstrukt Sachrechnen nahekommen, jedoch kann die inhaltlich größte Nähe zwischen KFT quantitativen Fähigkeiten und dem DEMAT vermutet werden.

Nachfolgend sollen die Residualkorrelationen der Subtestbeziehungen in der folgenden Grafik mit ihren Ergebnissen dargestellt werden. Die Subtestbeziehungen im Modell 6 unter Ermittlung kreuzverzögerter Effekte auf latenter Ebene erfolgt mit Angabe der Faktorladungen. Zunächst werden die Residualkorrelationen der KFT-Subtests auf latenter Ebene sind in der untenstehenden Abbildung veranschaulicht, während die Residualkorrelationen auf latenter Ebene für die mathematischen Subtests des DEMAT über die Zeit in der darauffolgenden Grafik gezeigt werden.

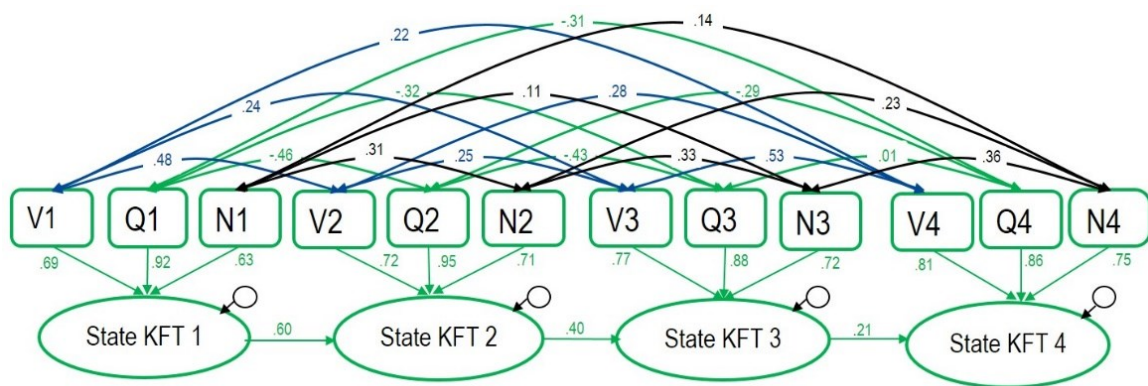


Abbildung 44 KFT-Subtestkorrelationen; Residualkorrelationen auf latenter Ebene.

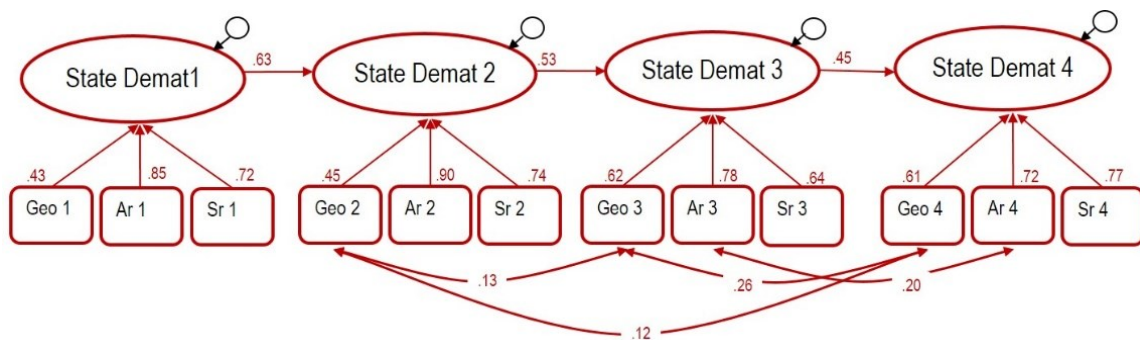


Abbildung 45 DEMAT-Subtestkorrelationen; Residualkorrelationen auf latenter Ebene;
Ar = Arithmetik; Geo = Geometrie; Sr = Sachrechnen.

Die komplette Darstellung der berücksichtigten Parameter im Crossed-Lagged-Modell 6 wird in der untenstehenden Abbildung veranschaulicht, wurde jedoch oben separat erläutert und wird an dieser Stelle nicht erneut ausgeführt.

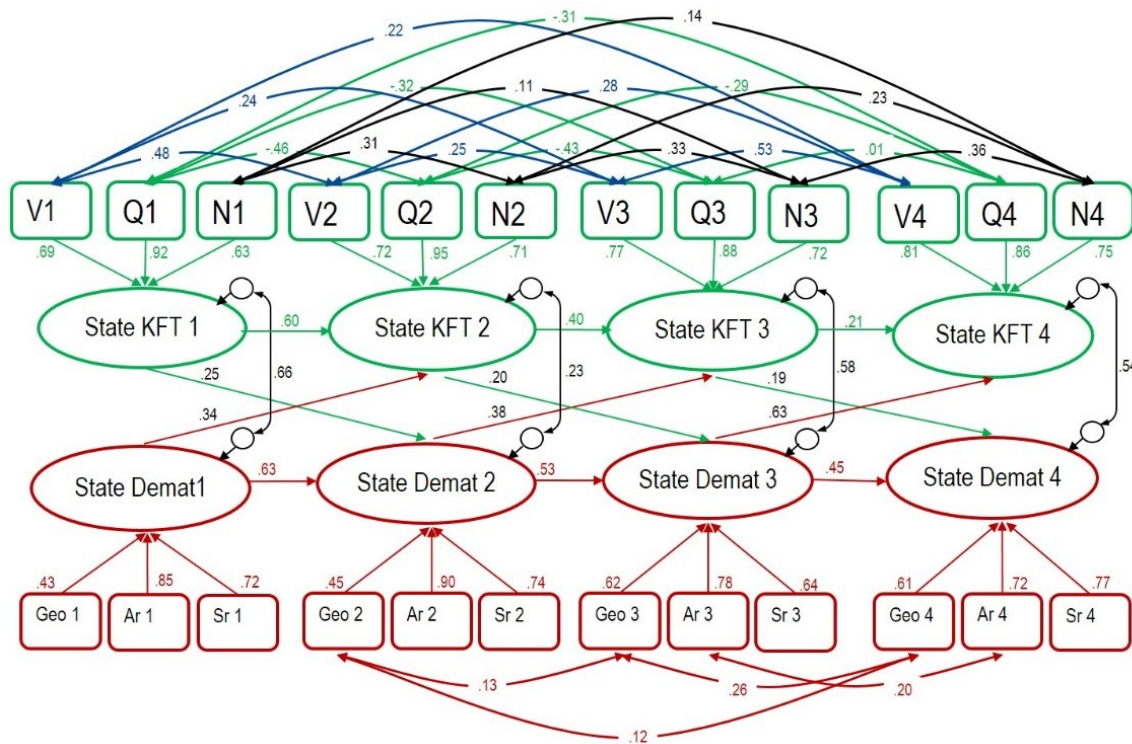


Abbildung 46 komplette Darstellung der berücksichtigten Parameter im Crossed-Lagged-Modell 6;
V = verbal; *Q* = quantitativ; *N* = nonverbal.

Im folgenden Abschnitt sollen die Modell-Fit-Informationen für die getesteten autoregressiven und Crossed-Lagged-Modelle kognitive Fähigkeiten und mathematische Schulleistungen über vier Zeitpunkte berichtet werden.

Betrachtung der Modellpassung

Die untenstehende Tabelle zeigt die Modellpassungswerte in der Übersicht mit allen sechs Modellen, die spezifiziert wurden. Dabei sind die ersten drei die autoregressiven Modelle zur Testung der Stabilität, während die Modelle 4,5 und 6 die Modellpassungswerte zur Ermittlung kreuzverzögerter Effekte zeigen. Zunächst wurden die autoregressiven Effekte des kognitiven Fähigkeitstests und des Testverfahrens zur Messung mathematischer Fähigkeiten im Modell der 1. Ordnung berechnet. Die Passung des autoregressiven Modells 1. Ordnung ist noch nicht optimal. Der χ^2 1348.298 mit 222 Freiheitsgraden. Der RMSEA liegt bei .06. Der CFI liegt bei .93 und der SRMR bei .68.

Im folgendem autoregressiven Modell *zweiter Ordnung* sollen neben den Effekten, die direkt auf den vorhergehenden Messzeitpunkt regredieren, auch Einflüsse auf zwei zu-

rückliegende Messzeitpunkte berechnet werden. Hier zeigt sich eine leicht bessere Modellpassung, die einen χ^2 von 1121.119 bei 218 Freiheitsgraden ausweist und einem besseren CFI von .94.

In der Konzeption der autoregressiven Effekte der *dritten Ordnung* sollen nunmehr die Regressionseffekte des ersten auf den dritten Messzeitpunkt einbezogen werden. Dieses Modell zeigt eine gute Passung mit einem CFI von .95, einem RMSEA von .05 und einem SRMR von .04, χ^2 von 1043.820 und 216 Freiheitsgraden.

In dem *vierten Modell* werden die kreuzverzögerten Effekte mit in die Modellkonzeption aufgenommen und dabei sollen zunächst die Regressionen des Konstrukts der mathematischen Leistungen vorhergesagt durch die kognitiven Fähigkeiten der vorhergehenden Messung untersucht werden. Der χ^2 -Wert weist 953.57 auf bei einer Anzahl an Freiheitsgraden von 213 ist die Modellgüte verbessert worden. Der RMSEA in Höhe von .04 und der SRMR von .03 und einem CFI von .97 lassen auf eine sehr gute Passung schließen.

Die autoregressiven Effekte der ersten und zweiten Ordnung sind in fast allen Bereichen signifikant und daher kann davon ausgegangen werden, dass ein bedeutsames Ausmaß an Stabilität in den interindividuellen Unterschieden in den untersuchten Merkmalen der Intelligenz und der mathematischen Fähigkeiten vorliegt. Lediglich zwischen den Pfaden des KFT 4 auf den KFT 2 und in den Pfaden des DEMAT 4 auf den DEMAT 1 lassen sich keine Signifikanzen feststellen. Die Untersuchung der Vorhersagekraft der kognitiven Fähigkeiten durch die Messung der mathematischen Leistungen kann im 5. Modell als Crossed-Lagged-Modell 1. Ordnung weiter spezifiziert werden. Dabei beschränkt sich die Konzeption der Regressionen zunächst nur auf die direkt vorangegangene Messung. Es zeigt sich im Ergebnis ein χ^2 von 659.598 und 210 Freiheitsgraden bei einem sehr guten RMSEA von .04 und einem SRMR von .03.

Die Beurteilung der Modellgüte des Modells unter Zulassung der kreuzverzögerten Effekte kann mit einem CFI von .98 als sehr gut bezeichnet werden. Das Modell 6 weist eine weitere Verbesserung mit einem χ^2 von 593.90 und 208 Freiheitsgraden, einem optimalen RMSEA von .03 und einem SRMR von .03. Der CFI ist mit einem Wert von .98 als sehr gut zu beurteilen.

Zusammenfassend zeigt der Vergleich der Modelle (siehe Tabelle unten) die Passungsverbesserung von dem lediglich autoregressiven Modell gegenüber dem Modell, wo einseitige Regressionen kognitiver Fähigkeiten auf mathematische Schulleistungen zugelassen wurden, bereits eine Verbesserung.

Tabelle 31

Passungswerte angewandter Modelle – autoregressiv (1 bis 3) und Crossed Lagged (4 bis 6)

Modell	1	2	3	4	5	6
χ^2	1348.30	1121.12	1043.82	953.50	659.60	593.90
df	222	218	216	213	210	208
<i>CFI</i>	.93	.94	.95	.95	.97	.98
<i>RMSEA</i>	.06	.05	.05	.05	.04	.03
<i>SRMR</i>	.68	.04	.04	.04	.03	.03
$\Delta\chi^2$		227.18	77.30	90.24	293.98	6.64
Δdf		4	2	3	3	2

χ^2 = Diskrepanz; $\Delta\chi^2$ = Differenz Diskrepanz; df = Freiheitsgrade; Δdf = Differenz Freiheitsgrade; *CFI* = Comparative Fit Index; *RMSEA* = Root Mean Square Error of Approximation (durchschnittliche Residuen); *SRMR* = standardisierte Residuen.

Das Modell mit nur regressiven Beziehungen kognitiver Fähigkeiten auf mathematische Schulleistungen zeigte bereits die Vorhersage der DEMAT Leistung durch den KFT vorheriger Messzeitpunkte. Es zeigte eine signifikant verbesserte Modellpassung, indem diese Pfade der regressiven Beziehungen hinzugenommen wurden.

Das Modell mit der Zulassung gegenseitiger kreuzverzögerter Regressionen (Modell 6), stellt das komplexe Modell mit kreuzverzögerten Pfaden dar. Es besitzt fünf Freiheitsgrade weniger, also 208 Freiheitsgrade im Vergleich zu 213 in dem Modell mit nur den regressiven Beziehungen kognitiver Fähigkeiten auf mathematische Schulleistungen, weil drei Pfade von DEMAT auf KFT hinzugenommen worden sind, und auch zwei Korrelationen zwischen Fehlern zugelassen wurden (zwischen den Residuen von KFT 2 und DEMAT 2 und KFT 3 und DEMAT 3). Dadurch konnte die Passung signifikant verbessert werden.

Es lässt sich schlussfolgern, dass sowohl die Reressionspfade von KFT auf DEMAT als auch die Regressionen des DEMAT auf den KFT signifikant sind. Es gibt also eine gegenseitige Beeinflussung mathematischer Schulleistungen und kognitiver Fähigkeiten. Die folgende Tabelle zeigt die regressiven Beziehungen aller sechs Modelle in der Übersicht.

Tabelle 32

Regressionen aller sechs Modelle – autoregressiv (1 bis 3) und Crossed Lagged (4 bis 6)

Pfad	1	2	3	4	5	6
<i>KFT 2 auf KFT 1</i>	.94	.91	.90	.89	.59	.60
<i>KFT 3 auf KFT 2</i>	.90	-.18	.05	.26	.43	.40
<i>KFT 4 auf KFT 3</i>	.89	.76	.13	.35	.42	.21
<i>KFT 3 auf KFT 1</i>		1.11	.87	.66	.11	.17
<i>KFT 4 auf KFT 2</i>		.17	-.24	-.08	-.07	.02
<i>KFT 4 auf KFT 1</i>			1.03	.67	.12	.16
<i>D 2 auf D 1</i>	.94	.89	.88	.35	.65	.63
<i>D 3 auf D 2</i>	.96	.09	.15	.49	.56	.53
<i>D 4 auf D 3</i>	.97	.78	.24	.42	.54	.45
<i>D 3 auf D 1</i>		.89	.81	.22	.20	.24
<i>D 4 auf D 2</i>		.21	.17	.27	.20	.28
<i>D 4 auf D 1</i>			.58	.05	.06	.10
<i>KFT 2 auf D 1</i>					.37	.34
<i>KFT 3 auf D 2</i>					.44	.38
<i>KFT 4 auf D 3</i>					.53	.63
<i>D 2 auf KFT 1</i>				.54	.26	.25
<i>D 3 auf KFT 2</i>				.29	.24	.20
<i>D 4 auf KFT 3</i>				.29	.22	.20

In der Tabelle sind die Regressionen aller sechs Modelle (1, 2, 3 = autoregressiv) und (4, 5, 6 = crossed-lagged) aufgeführt.

Auch die Kontrolle der Mehrebenenstruktur in den Modellen zur Ermittlung kreuzverzögerter Effekte erfolgte jeweils modellspezifisch, wie die untenstehende Tabelle zeigt. Die Passungswerte der eingesetzten Modelle mit und ohne Mehrebenenkontrolle sind in der Tabellenübersicht aufgeführt.

Tabelle 33

Passungswerte angewandter Modelle – Passungswerte angewandter Modelle zur Prüfung der kreuzverzögerten Effekte (Modelle 4-6) mit und ohne Mehrebenenkontrolle

Modell	4 ohne MEK	4 mit MEK	5 ohne MEK	5 mit MEK	6 ohne MEK	6 mit MEK
χ^2	953.57	865.99	659.60	593.44	593.90	533.37
<i>df</i>	213	213	210	210	208	208
<i>CFI</i>	.95	.96	.97	.98	.98	.98
<i>RMSEA</i>	.05	.04	.04	.03	.03	.03
<i>SRMR</i>	.04	.04	.03	.03	.03	.03

MEK = Mehrebenenkontrolle; χ^2 = Diskrepanzf; *df* = Freiheitsgrade; *CFI* = Comparative Fit Index; *RMSEA* = Root Mean Square Error of Approximation (durchschnittliche Residuen); *SRMR* = standardisierte Residuen.

Dabei ist festzustellen, dass die erhobene Stichprobe zwar geclustert vorlag, aber kein wesentlicher Unterschied festgestellt werden konnte. Die Kontrolle der Clusterung auf Klassenebene zu untersuchen, wäre eine empfehlenswerte Frage zukünftiger Untersuchungen, um die Beziehungen zwischen mathematischen Schulleistungen und kognitiven Fähigkeiten in Abhängigkeit der Klassenstrukturen zu untersuchen.

Im folgenden Abschnitt sollen nun die Ergebnisse der Untersuchungen kurz zusammengefasst werden.

4.5.4 Kreuzverzögerte Effekte – Zusammenfassung und Diskussion der Ergebnisse

Um die kreuzverzögerten Effekte kognitiver Fähigkeiten und mathematischer Schulleistungen zu ermitteln, wurde zunächst das autoregressive Modell mit der besten Modellpassung als Basismodell für den Aufbau der kreuzverzögerten Effekte zugrunde gelegt. Das als Modell 6 konzipierte Modell mit der Spezifikation der kreuzverzögerten Effekte konnte anhand der Faktorladungen der Subtests den größten Aufklärungsbeitrag leisten. Dieses Modell erwies sich als das geeignetste und bestätigt die autoregressiven Einflüsse der jeweiligen Einflüsse untereinander über die vier Messzeitpunkte, aber, und das sei an dieser Stelle besonders betont, die gegenseitige Beeinflussung mathematischer Schulleistungen und kognitiver Fähigkeiten. Der Einfluss mathematischer Schulleistungen als Prädiktoren kognitiver Fähigkeiten war dabei höher als in der Gegenrichtung. Die bessere Eignung ließ sich anhand der quadrierten Faktorladung erkennen, die ein Qualitätskriterium für den Aufklärungsgrad darstellt. Die bessere Eignung erwies sich auch durch die gegenüber dem Vormodell zugelassenen Residualkorrelationen, die dort nicht kovariieren durften. Damit war die Vorhersage des KFT aufgrund der DEMAT Werte zuvor nicht zugelassen. In dem komplexesten Modell konnten also mehr erklärende Ergebnisse beigetragen werden, als im vorherigen Modell, bei dennoch guter Modellpassung. Einige negative Residualkorrelationen sind zu akzeptieren, jedoch ist wiederum positiv, dass die Faktorkorrelationen bzw. Faktorladungen äußerst zufriedenstellend sind. Zudem sind die Residualkorrelationen nicht sehr verschieden, z. B. innerhalb der verbalen Subskala des kognitiven Fähigkeitstests mit einer Höhe von .69 und .81, besteht eine Differenz von .12. Ab .20 könnte von großen Differenzen ausgegangen werden. Die Faktorkorrelationen sind dagegen positiv, wenn auch gering.

Der Zusammenhang zwischen Mathematikleistungen und kognitiven Fähigkeiten ist über den Betrachtungszeitraum positiv. Die Ergebnisse der Crossed-Lagged-Effekte unter Hinzuziehung der Residualkovarianzen (DEMAT vorhergesagt durch KFT, KFT vorhergesagt durch DEMAT) zeigen bei der Betrachtung die zunehmende Vorhersagekraft über die Jahre bei sehr guten Modellfitinformationen ($\chi^2 = 593.956$; $df = 208$; $RMSEA = .03$; $SRMR = .03$; $CFI = .98$). Die quadrierte Faktorladung der latenten Variablen fällt bei der Betrachtung des KFT höher aus als in den vorherigen Modellspezifikationen. Damit zeigen die Reliabilitätswerte eine hohe bis sehr hohe Ausprägung und zeugen von einer hohen zeitlichen Stabilität der Merkmale. Auch für den DEMAT zeigen sich Reliabilitätswerte zwischen .67

und .91. Damit kann das latente Modell deutlich präziser die Varianz des zweiten, dritten und vierten Messzeitpunktes vorhersagen.

Es wurde erwartet, dass kognitive Fähigkeiten die mathematischen Schulleistungen über die vier betrachteten Messzeitpunkte zu späteren Zeitpunkten beeinflussen. Diese Hypothese kann mit den vorliegenden Ergebnissen bestätigt werden.

Es wurde in Hypothese 4 eine Wirkung früherer mathematischer Schulleistungen auf spätere kognitive Leistungen über die vier betrachteten Messzeitpunkte erwartet. Diese Hypothese kann ebenfalls angenommen werden.

Es wurde davon ausgegangen, dass sich kognitive Leistungen und mathematische Schulleistungen über die vier betrachteten Messzeitpunkte gegenseitig beeinflussen. Diese Hypothese kann ebenso angenommen werden.

Es wurde zudem erwartet, dass sich diese Effekte bereits von der ersten Klassenstufe an mit Wirkung auf die zweite Klassenstufe zeigen. Diese Hypothese ließ sich ebenfalls bestätigen.

Es wurde angenommen, dass sich die Effekte eines Konstrukts auf das jeweils andere im Laufe der Grundschulzeit verstärken, sodass erwartet wurde, dass sich die größten gegenseitigen Effekte von der dritten zur vierten Klassenstufe des Gegenkonstrukts zeigen. Diese Hypothese kann ebenso angenommen werden.

Das primäre Ziel der vorliegenden Untersuchungen war, die Stärke und Richtung der Beziehungen zwischen kognitiven Fähigkeiten und mathematischen Schulleistungen zu bestimmen. Die Analysen mittels latenter Strukturgleichungsmodelle über die vier Messzeitpunkte bestätigen gegenseitige Einflüsse beider Konstrukte. Wesentlichen Einfluss auf die Qualität der Ergebnisse hat dabei die Spezifikation anhand zweier Testverfahren, die sowohl für die kognitiven Fähigkeiten auf bereichsspezifischen Messungen basierte als auch die stoffgebietsspezifische Unterscheidung der mathematischen Schulleistungen der DEMAT-Testreihe vornimmt. Es stand folglich eine große Anzahl Items für die jeweiligen Subskalen zur Verfügung, sodass der Spezifikation der Modelle eine hohe Zuverlässigkeit zugerechnet werden kann. Berichte über die Stärke und Richtung der Zusammenhänge beider Konstrukte lagen für das Grundschulalter über vier Messzeitpunkte in dieser Form noch nicht vor.

Die ermittelten Ergebnisse bestätigen den Einfluss kognitiver Fähigkeiten auf schulische Mathematikleistungen, aber auch den Einfluss von mathematischen Schulleistungen auf kognitive Fähigkeiten über die Zeit. Dabei kann zusätzlich hervorgehoben werden, dass der zunehmende Einfluss von mathematischen Schulleistungen auf spätere kognitive Fähigkeiten mit dem Vorwissen erklärt werden kann. Die ermittelten Zusammenhänge stimmen somit mit vorherigen Studien überein, die nachweisen konnten, dass das schulisch erworbene Wissen sowie das in der Expertisetheorie diskutierte relevante Vorwissen die späteren kognitiven Fähigkeiten vorhersagen können. Dabei bleibt der Einfluss des jeweils selben Konstrukts über die Zeit stabil, sodass geschlussfolgert werden kann, dass kognitive Fähigkeiten zugleich spätere kognitive Fähigkeiten beeinflussen, während auch mathematische Schulleistungen über die vier Messzeitpunkte einen bedeutenden Einfluss beibehalten. Die gegenseitigen Effekte von kognitiven Fähigkeiten und mathematischen Schulleistungen, deuten darauf hin, dass kognitive Fähigkeiten durch schulisches Wissen beeinflusst werden und damit ein kompensierender Effekt vermutet werden kann.

Der Einfluss kognitiver Fähigkeiten als Prädiktor mathematischer Schulleistungen kann gleichwohl bestätigt werden, sodass auch von einem kumulativen Einfluss ausgegangen werden kann. Der relevante Einfluss kognitiver Fähigkeiten auf die mathematischen Schulleistungen kann mit den für die erfolgreiche Bewältigung von Herausforderungen aus beiden Bereichen notwendigen Problemlösefähigkeiten gut erklärt werden. Mathematische Leistungen erfordern kognitive Fähigkeiten wie Problemlösefähigkeit und schlussfolgerndes Denken (Floyd, Evans und McGrew, 2003). Weiterhin machen die Ergebnisse deutlich, dass es lohnenswert erscheint, ein hierarchisch konzeptionalisiertes domänenspezifisches Theorieverständnis der Testverfahren zu bevorzugen, da deutlich wurde, dass auf latenter Ebene die Beziehungen entsprechend der Subskalen deutlich voneinander differenzierbar sind.

Die jeweiligen Ausprägungen der Leistungen in den Bereichen lassen auf Einflüsse des Wissens und Vorwissens sowohl auf Seiten der kognitiven Fähigkeiten als auch auf Seiten der mathematischen Schulleistungen schließen. Damit stehen die Ergebnisse im Einklang mit den Positionen, die von einer gemäßigten Expertisetheorie ausgehen (Perleth, 1997) als auch den Autoren, die bislang den Einfluss von Vorläuferfertigkeiten auf Seiten des schulisch und vorschulisch erwerbbareren Wissens sehen (Weinert & Helmke, 2003; Krajewski & Schneider, 2006; Krajewski, 2008; Helmke & Weinert, 1997).

5 Gesamtdiskussion, Implikationen und Perspektiven

Die Ergebnisse der vorliegenden Arbeit sollen abschließend einer kritischen Prüfung unterzogen werden.

Mit einem kurzen selbstkritischen Blick auf die Arbeit soll begonnen werden. Einleitend wurde bereits auf die Schwierigkeiten bei der Definition der für die Arbeit leitenden Begriffe hingewiesen. So führte im Rahmen dieser Arbeit die mangelnde Präzision bei den Definitionen von Vorwissen, das entweder schulisch erworben als Bestandteil der mathematischen Schulleistungen oder auf der anderen Seite als Vorwissensanteil im Sinne kognitiver Fähigkeiten den Messinstrumenten zugrunde liegt, zu einer ungenauen Unterscheidung zwischen Wissensstrukturen als Vorkenntnissen und kognitiven Fähigkeiten als Intelligenz. Hier muss von einem Überschneidungsbereich ausgegangen werden, der so nicht sauber differenziert worden ist. Die problematische Unterscheidung zwischen den Begriffen Intelligenz und schulische Leistung erwähnten bereits (Cahen & Cohan, 1986). Es ist daher zu empfehlen, in zukünftigen Untersuchungen die Messinstrumente dahingehend zu prüfen, inwiefern sie eine präzise theoretische Trennung der angenommenen Konstrukte berücksichtigen. Es hat sich gezeigt, dass die stoffgebietsspezifische Unterscheidung mathematischer Schulleistungen ebenso notwendig ist, wie die konzeptionelle Unterscheidung kognitiver Fähigkeiten nach ihren Domänen und ihrer hierarchischen Struktur. Es ist daher sehr zu empfehlen, die Messinstrumente für längsschnittliche Untersuchungen hinsichtlich ihrer Konstruktkonsistenz zu überprüfen. Dann sind noch präzisere längsschnittliche Aussagen möglich, die wiederum bedeutsam für die weiterführende theoretische Diskussion und die pädagogische Praxis sind.

Die vorliegende Arbeit kann allerdings trotz dieser kleinen methodischen Einschränkung wichtige Impulse sowohl für die Theorie als auch die Praxis geben. Weiterführende Forschungsfragen und Handlungsoptionen für die Praxis werden daher im Folgenden zur Diskussion gestellt.

Für die pädagogisch-psychologische Forschung können die vorliegenden Resultate für mehrere Forschungsteilbereiche interessante Ansätze bieten.

Hinsichtlich der theoretisch intensiv geführten Debatte um die Frage, ob kognitive Leistungen und mathematische Schulleistungen womöglich identische Konstrukte erfassen, bzw. inwiefern es sich um überlappende Konstrukte handelt oder ob diese klar voneinander zu unterscheiden sind, konnte die strukturelle Unabhängigkeit beider Konstrukte überzeu-

gend nachgewiesen werden. Anhand der Ergebnisse dieser Arbeit können theoretische Ansichten empirisch unterstützt werden, welche die klare Differenziertheit beider Konstrukte vertreten. Jedoch lenkt dieses Ergebnis den Blick auf die Problematik der Vielzahl gleichzeitig existierender Definitionen von kognitiven Fähigkeiten. Zudem wird deutlich, dass in der Praxis der diagnostischen Anwendung häufig nicht klar abgegrenzt wird, wie kognitiv verstandenes Vorwissen sich von schulisch erworbenen Vorwissen abgrenzen lässt, obwohl empirisch belegte Ansätze existieren. Eine Konsequenz wäre die Umsetzung bereichsspezifischer Erfassung der Konstrukte. Mit der Replikation der Ergebnisse dieser Arbeit und dem erneuten Nachweis der faktoriellen Differenziertheit auch in anderen Verfahren könnten jene Forschungspositionen gestützt werden, die sich für den Einsatz von Messverfahren aussprechen, die konzeptionelle gezielt bereichsspezifische Merkmale erfassen.

Aussichtsreich ist, wenn sich diese Ergebnisse nicht nur für das Grundschulalter, sondern auch für das Sekundarstufenalter bzw. von der Grundschulzeit über einen längeren Zeitraum bis in das Sekundarstufenalter replizieren lassen. Dabei sind Analysen dann besonders aussagekräftig, wenn für jeden Subtest möglichst zahlreiche Items zur Verfügung stehen, denn nur so lassen sich diese explizieren, da sich bei einer geringeren Anzahl keine reliablen Indikatoren nachweisen lassen können. Die Aussagekraft der hier berichteten empirischen Befunde ließe sich nochmals erhöhen, wenn ein höherer Invarianzgrad der Messungen über die eingesetzten bereichsspezifischen Testverfahren hätte erreicht werden können. Unter der gegebenen partiellen Invarianz der mathematischen Testreihe mussten somit einige Items aus der Analyse ausgeschlossen werden, was die Aussagekraft (wenn auch nur gering) schmälerte.

Abschließend soll betont werden, dass mittels der Spezifikation von Latent State Modellen eine klare Trennung der Konstrukte kognitive Fähigkeiten und mathematische Schulleistungen über die vier Messzeitpunkte nachgewiesen werden konnte und damit empirisch belegbar ist, dass mathematische Schulleistungen nicht identisch mit kognitiven Fähigkeiten sind.

Die Ergebnisse der Prüfung der zweiten Fragestellung hinsichtlich der Konstruktkonsistenz des Testverfahrens von kognitiven Fähigkeiten auf der einen Seite und der Konstruktkonsistenz der Testverfahren zur Erfassung mathematischer Schulleistungen auf der anderen Seite zeigten zweierlei Ergebnisse: Die Konstruktkonsistenz der Testverfahren aus der KFT-Testreihe zeigte invariantere (und damit bessere) Ergebnisse als die Invarianzprüfung der Testverfahren mathematischer Schulleistungen. Die besseren Resultate der Test-

reihe zur Erfassung kognitiver Fähigkeiten könnten darin begründet sein, dass diesen Testverfahren nur zwei Testadaptionen zugrunde lagen, sodass das Konstruktverständnis geringeren Veränderungen aufgrund der Anpassungskonzeptionen für die jeweiligen Klassenstufen unterlag. Jedoch zeigten sich in der Einzelbetrachtung der Subskalen auch Bereiche, in denen die Invarianz weniger gut nachweisbar war. Der metrische bzw. skalare Invarianzgrad konnte also nicht erreicht werden, sodass weitere Forschung im Bereich der Kriteriumsvalidität der Testreihe aussichtsreich wäre, um einen höheren Invarianzgrad zu erreichen.

Um die Konstruktkonsistenz für Längsschnittstudien geeigneter erscheinen zu lassen, ist zu empfehlen, diese Konstruktkonsistenz der jeweiligen Items in den Subskalen der Testverfahren über die vier Klassenstufen hinweg zu überprüfen. Damit verringert sich die Gefahr einer Verzerrung von Ergebnissen aufgrund von kumulativen Messfehlern, die durch Konstruktionstimmigkeiten verursacht worden sind. Letztlich wird die Aussagekraft von Ergebnissen so wesentlich präziser. Weitere Forschung im Bereich der Kriteriumsvalidität der Testreihe ist hier aussichtsreich, um einen höheren Invarianzgrad zu erreichen.

Sofern Leistungen im zeitlichen Verlauf im Vordergrund stehen, hängt die Aussagekraft von Resultaten erheblich von der Konstruktkonsistenz der in dem betrachteten Zeitraum eingesetzten Messinstrumente ab, auf die es besonderes Augenmerk zu legen gilt.

Die Erklärung der nur partiell nachgewiesenen Invarianz mathematischer Schulleistungen ist ebenfalls durch Adaptionen der Testverfahren von der ersten zur zweiten, zur dritten und zur vierten Klassenstufe zu erklären. Das explizite Anliegen der Testautoren der mathematischen Testverfahren, ein stark am Curriculum orientiertes Kompetenzerfassungsinstrument für die jeweiligen Klassenstufen und somit vor allem für Jahrgangsorientierungen zu konstruieren, ist hier ursächlich zu nennen. Um auch für die Stoffgebiete Geometrie und Sachrechnen invariante Ergebnisse zu erzielen, ist zu empfehlen, die Konstruktkonsistenz der entsprechenden Items dieser vier Testadaptionen einer kritischen Prüfung zu unterziehen. Die vorhandenen Testverfahren sind zwar jeweils für die entsprechende Kompetenzmessung des jahrgangsspezifischen Curriculums sehr gut geeignet, doch nur bedingt geeignet, die Entwicklung der Mathematikleistungen von Grundschulkindern über einen längeren Zeitraum präzise darzustellen. Eine Prüfung der Testreihe unter dem Aspekt der Konstruktkonsistenz über die ersten vier Klassenstufen hinweg könnte diese Testreihe zum favorisierten Instrument für Längsschnittstudien, sowohl für die pädagogische Praxis in der Lernerfolgskontrolle als auch für die wissenschaftliche Forschung machen.

Sofern im Rahmen der Evaluationsforschung beabsichtigt wird, längsschnittliche Fragestellungen zu beantworten, sollte die Konstruktkonsistenz der eingesetzten Testverfahren bereits vor dem Einsatz berücksichtigt werden, da sie wesentlich die Höhe der Messfehler bestimmt. Dies gilt auch für die Bewertung vormaliger Längsschnittdaten zu diesem Thema. Im Rahmen längsschnittlicher psychologischer Forschung wäre die Prüfung der Konstruktkonsistenz sicherlich ein Qualitätsmerkmal.

Um zuverlässige Aussagen aus Längsschnittdaten abzuleiten, sollten Längsschnittstudien situationsspezifische bzw. messfehlerbedingte Einflüsse berücksichtigen, deren unerwünschte Kumulationen bei Unterlassung der Kontrolle über einen längeren Untersuchungszeitraum zu verzerrten Schlussfolgerungen führen können.

Hinsichtlich der Einordnung der Ergebnisse der vierten Fragestellung in den bisherigen theoretischen Kenntnisstand kann folgendes Fazit gezogen werden: Die Resultate der kreuzverzögerten Effekte stützen deutlich die theoretische Ansicht, dass im Grundschulalter keine einseitige Beeinflussung durch kognitive Fähigkeiten auf schulische Mathematikleistungen, sondern eine wechselseitige Beeinflussung auch von Mathematikleistungen auf kognitive Fähigkeiten empirisch zu beobachten ist. Die Ergebnisse zeigen, dass Wissen und Vorwissen eine dominierende Rolle hinsichtlich allgemeiner und spezifischer kognitiver Leistungen in den Folgejahren spielen. Damit können diese Resultate eine bedeutende Relevanz für die pädagogische Praxis besitzen, beispielsweise hinsichtlich des Transfers von Lerntrainings, und erweitern damit den theoretischen Kenntnisstand aufgrund der nunmehr gelungenen empirischen Belegbarkeit der Einflüsse mathematischen Wissens auf kognitive Fähigkeiten. Bereits Hasselhorn und Hager (2008) oder Perleth (1997) und Klauer (2011) waren von der Bedeutung von Wissenselementen im Rahmen von Lerninterventionsmaßnahmen überzeugt. Die theoretisch geführten Debatten, welche die Beeinflussung von kognitiven Fähigkeiten durch (Vor-)Wissenselemente betonen, siehe Campione und Brown (1978), Borkowski und Beck (1986) oder auch Schneider (1992), auch Jäger (1982, 1984) und Perleth (1997), welche den tragenden Einfluss von Wissenserwerbskomponenten im Rahmen der Leistungserbringung deutlich machen, können mit den Ergebnissen dieser Arbeit unterstützt werden.

Die hier vorliegenden Resultate können die bislang rare Befundlage um aufschlussreiche Forschungsergebnisse ergänzen, welche die gegenseitige Einflussrichtung von Schulleistungsmerkmalen und kognitiven Fähigkeiten integrieren und sollten bei der Konzeption von Folgestudien einbezogen werden.

Die Stabilität der interindividuellen Unterschiede ist sowohl für die kognitiven Fähigkeiten als auch für die mathematischen Schulleistungen relativ hoch, wenn sich auch bei einzelnen Kindern Verschiebungen in den Rangreihen sowohl nach oben als auch nach unten finden. Die in Fragestellung drei thematisierte Frage der gegenseitigen Wechselwirkungen sowie die in der theoretischen Forschung diskutierten kumulativen Effekte um den „Matthäus-Effekt“ bzw. kompensative Effekte, könnten eine Erklärung für die interindividuell vorhandenen Schwankungen sein. Die moderierenden Faktoren für diese Kinder zu ermitteln, wäre eine forschungsrelevante Frage, die sich mit dem vorliegenden Datensatz beantworten ließe.

Auch die mathematischen Schulleistungen zeigen eine hohe interindividuelle Stabilität. Da in dieser Arbeit Kinder im Grundschulalter im Mittelpunkt der Untersuchung standen, kann die weiterführende Forschungsfrage von Interesse sein, welche Stabilitätswerte für ältere Kinder bzw. Jugendliche unter den Bedingungen der Berücksichtigung von Messfehlern auf latenter Ebene zu ermitteln sind.

Die nachweisbaren kreuzverzögerten Effekte lassen den Schluss zu, dass begründet angenommen werden kann, dass nicht nur kognitive Fähigkeiten spätere Mathematikleistungen beeinflussen, sondern dass auch (vor)wissensbasierte mathematische Leistungen einen deutlichen Einfluss auf spätere kognitive Fähigkeiten besitzen, deren Einfluss im Laufe der Grundschulzeit sogar zunimmt. Dieses Ergebnis kann vor allem pädagogische Relevanz haben, denn es stellt in Aussicht, dass Interventionen, die an der Organisation (bzw. Wissenserwerbsprozessen) von Vor-(Wissen) ansetzen, nicht nur erfolgreich die Leistungen des jeweiligen Bereiches verbessern können, sondern auch Transferwirkungen und damit das Kompensationspotential hinsichtlich allgemeiner Denkfähigkeiten beeinflussen können. Diese Ergebnisse lassen sich über den gesamten Grundschulzeitraum hinweg nachweisen und nehmen im Laufe der Grundschulzeit sogar zu.

Ein wesentlicher Aspekt ist auch das Ergebnis, dass kognitive Fähigkeiten als Prädiktoren schulischer Mathematikleistungen gelten, doch diese Prädiktionskraft nicht einseitig zu verstehen ist, sondern in einer wechselseitigen Beziehung über die Zeit besteht.

Neben der pädagogisch-psychologischen Forschung könnte auch die methodisch orientierte psychologische Forschung von den Ergebnissen dieser Arbeit profitieren. So könnten noch präzisere Wirkbeziehungen hinsichtlich schul- und klassenspezifischer Unterschiede zu den Wechselwirkungen kognitiver Fähigkeiten und mathematischer Schulleistungen offengelegt werden. Da sich die Kinder, deren Testergebnisse die Grundlage für diese Arbeit bilden, in verschiedenen Schulen befanden, lag eine schulspezifische Clusterung der Datenstruktur vor. Die Varianzunterschiede zwischen den Schulen fallen größer

aus als Varianzunterschiede innerhalb einer Schule. Die somit vorliegende Mehrebenenstruktur der Stichprobe wurde in dieser Analyse kontrolliert, allerdings nicht vollständig in den Analysen berücksichtigt, sodass hier eine noch höhere Genauigkeit zu erzielen ist. Ein anderer interessanter Ansatz, der Aufschluss zu den individuellen Entwicklungen kognitiver Fähigkeiten und mathematischer Schulleistungen geben könnte, wäre anhand der Spezifikation von Wachstumskurvenmodellen realisierbar. Die Datengrundlage erlaubt eine detaillierte Analyse der Mehrebenenstrukturen innerhalb der 75 Schulklassen und könnte Gegenstand zukünftiger Analysen sein.

Doch die Ergebnisse der vorliegenden Arbeit können nicht nur vielversprechende Anregungen für die Forschungsdebatte liefern, sondern auch Handlungsimpulse für die Praxis anregen.

Aufgrund der gezeigten zunehmenden Wirkung von mathematischen Schulleistungen auf kognitive Fähigkeiten könnten vielversprechende Ableitungen für die schulpädagogische Praxis getroffen werden, welche den Einfluss von wissensfördernden Ansätzen auf kognitive Leistungen stärker betonen. Insbesondere die kompensativen Effekte schulischer Leistungen auf kognitive Fähigkeiten, die wiederum ihre Wirkung auf nachfolgende schulischen Leistungen zeigen, können hinsichtlich der interventionsorientierter Ansätze interessante Themen eröffnen.

Dabei ist es relevant, auf welchem Konzept die eingesetzten Testverfahren basieren. So sind bereichsspezifische kognitive Testverfahren zum einen in der Lage, die Domänenspezifität des kognitiven Fähigkeitsspektrums abzubilden. Sie können somit im Sinne eines Profils die individuellen Stärken und Schwächen der Person aufdecken und pädagogisch nutzbar machen. Zum anderen ist es wichtig, dass bereichsspezifische Testverfahren eingesetzt werden, die auch explizit Wissensanteile und wesentliche Aspekte des Vorwissens erheben. Der Einfluss des Vorwissens ist entscheidend für die weitere Bildungslaufbahn und sollte nicht in einem nur fluide, allgemeine Denkfähigkeiten erfassenden Testverfahren aufgehen. Dennoch ist auch der Erfassung allgemeiner, sprachfreier Intelligenzanteile weiterhin Bedeutung beizumessen, da sie Aussagen zur Denkflüssigkeit (im Sinne Cattells) und Vernetzung im Kontext unbekannter Probleme treffen kann.

Auch die standardisierte Erfassung von Schulleistungen mittels Schulleistungstestverfahren ist der Erfassung mittels Schulnoten vorzuziehen, da Schulleistungstests den standardisierten Gütekriterien entsprechen und darüber hinaus verschiedene Domänen unterschiedlicher Wissensstrukturen sichtbar machen. Also werden nicht nur allgemeine Mathematikleistungen erfassbar, sondern stoffgebietsspezifische Differenzen im Niveau können

den individuellen pädagogischen Nutzen unterstreichen. Auch im Rahmen individueller Beratung ist die Abbildung eines standardisierten Leistungsprofils ein transparentes Diagnoseinstrument, welches sich darüber hinaus für die Dokumentation im Prozessverlauf pädagogischer Interventionen eignet.

Schlussfolgernd kann resümiert werden, dass kognitive Fähigkeiten und schulische Leistungen multidimensional betrachtet werden sollten, gerade vor dem Hintergrund der Bewältigung von Lernschwierigkeiten.

Zusammenfassend ist festzuhalten, dass die vorliegende Arbeit im Rahmen der offenen Forschungsfragen zu den Zusammenhängen zwischen kognitiven Fähigkeiten und Schulleistungen einen wichtigen und innovativen Beitrag leisten konnte. Sie hat aber auch gezeigt, dass noch viele Fragen offenbleiben. Der Forschungsbedarf im Spannungsfeld zwischen kognitiven Fähigkeiten und Schulleistungen ist weiterhin hoch.

6 Verzeichnisse

6.1 Literaturverzeichnis

- Abou-Koura, K. (2005). Reliabilitäts- und Validitätsuntersuchungen zum neuen Kognitiven Fähigkeitstest für die Primarstufe (KFT 1-2 R) unter besonderer Berücksichtigung von Kindern mit Hochbegabungen und Lernbehinderungen (Unveröffentlichte Dissertation). Universität Rostock.
- Abou-Koura, K. & Perleth, Ch. (2005). *Kognitiver Fähigkeitstest (KFT 1-2R)* (Unveröffentlichte Revision). Universität Rostock.
- Ackerman, P. L. (1986). Individual differences in information processing: An investigation of intellectual abilities and task performance during practice. *Intelligence*, 10, 101–139.
- Ackerman, P.L. (1989). Individual differences and skill acquisition. In P.L. Ackerman, R. J. Sternberg, & R. Glaser (Eds.), *Learning and individual differences* (pp. 164–217). New York: W.H. Freeman and Company.
- Ackerman, P.L. (2003). Cognitive ability and non-ability. *Educational Researcher*, 32 (8), 15–20.
- Ackerman, P.L. (2011). Intelligence and expertise. In R.J. Sternberg, & S.B. Kaufman (Eds.), *The Cambridge handbook of intelligence* (pp. 847–860). Cambridge University Press.
- Ackerman, P.L., Beier, M.E. & Boyle, M.O. (2005). Working memory and intelligence: The same or different constructs? *Psychological Bulletin*, 131 (1), 30–60.
- Aebli, H. (1976). *Grundformen des Lehrens: Eine allgemeine Didaktik auf kognitionspsychologischer Grundlage* (9. Aufl.). Stuttgart: Klett.
- Agostino, A., Johnson, J. & Pascual-Leone, J. (2010). Executive functions underlying multiplicative reasoning. Problem type matters. *Journal of Experimental Child Psychology*, 105 (4), 286–305.
- Akukwe, B. & Schroeders, U. (2016). Socio-economic, cultural, social, and cognitive aspects of family background and the biology competency of ninth-graders in Germany. *Learning and Individual Differences*, 45, 185–192.
- Albrecht, L. (1926). *Das Neue Testament* (5. Auflage). Gotha: Verlag der Evangelischen Buchhandlung: Ott.
- Alexander, K.L., Entwisle, D.R. & Olson, L.S. (2001). Schools, achievement, and inequality. A seasonal perspective. *Educational Evaluation and Policy Analysis*, 23, 171–191.
- Alfonso, V.C., Flanagan, D. P. & Radwan, S. (2005). The impact of the Cattell-Horn-Carroll theory on test development and interpretation of cognitive and academic abilities. In D.P. Flanagan & P. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (2nd edition, pp. 185–202). New York: Guilford Press.
- Anderson, J.R. (2007). *Kognitive Psychologie* (6. Aufl.). Berlin: Springer-Verlag.
- Andersson, U. & Lyxell, B. (2007). Working memory deficit in children with mathematical difficulties. A general or specific deficit? *Journal of Experimental Child Psychology*, 96 (3), 197–228.

- Ashcraft, M.H. & Fiermann, B.A. (1982). Mental addition in third, fourth and sixth graders. *Journal of Experimental Child Psychology*, 33 (2), 216–234.
- Baddeley, A.D. (1986). *Working memory*. New York: Oxford University Press.
- Baddeley, A.D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Science*, 4, 417–423.
- Baddeley, A.D. (2006). Working memory: An overview. In S.J. Pickering (Ed.), *Working memory and education* (pp. 1 – 31). San Diego, CA: Academic Press.
- Bast, J. & Reitsma, P. (1998). Analyzing the development of individual differences in terms of Matthew effects in reading: Results from a Dutch longitudinal study. *Developmental Psychology*, 34, 1373–1399.
- Baumert, J., Blum, W., Brunner, M., Dubberke, T., Jordan, A., Klusmann, U., et al. (2009). *Professionswissen von Lehrkräften, kognitiv aktivierender Mathematikunterricht und die Entwicklung von mathematischer Kompetenz (COACTIV): Dokumentation der Erhebungsinstrumente (Materialien aus der Bildungsforschung Nr. 83)*. Berlin: Max-Planck-Institut für Bildungsforschung.
- Baumert, J., Brunner, M., Lüdtke, O. & Trautwein, U. (2007). Was messen internationale Schulleistungsstudien? – Resultate kumulativer Wissenserwerbsprozesse. *Psychologische Rundschau*, 58 (2), 118–145.
- Baumert, J., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, W., Stanat, P., Tillmann K.-J. & Weiß, M. (Hrsg.) (2001). *Pisa 2000: Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich*. Opladen: Leske + Budrich.
- Baumert, J. & Lehmann, R. (1997). *TIMSS – Mathematischnaturwissenschaftlicher Unterricht im internationalen Vergleich*. Opladen: Leske Budrich.
- Baumert, J., Lüdtke, O., Trautwein, U. & Brunner, M. (2009). Large-scale student assessment studies measure the results of processes of knowledge acquisition: Evidence in support of the distinction between intelligence and student achievement. *Educational Research Review*, 4, 165–176.
- Berger, P.A. & Kahlert, H. (Hrsg.) (2013). *Institutionalisierte Ungleichheiten – Wie das Bildungssystem Chancen blockiert* (3. Aufl.). Weinheim: Juventa.
- Binet, A. & Simon, T. (1916). *The development of intelligence in children* (E. S. Kite, Trans.). Baltimore, MD: Williams & Wilkins.
- Bloom, B.S. (1971). *Stabilität und Veränderung menschlicher Merkmale*. Weinheim: Verlag Julius Beltz.
- Blum, W. & Krauss, S. (2007, November). *The professional knowledge of German secondary mathematics teachers: Investigations in the context of the COACTIV Project*. Vortrag im Mathematischen Forschungsinstitut Oberwolfach.
- Blumenthal, Y., Hartke, B. & Koch, K. (2009). Zur Lernausgangslage von Kindern mit besonderem Förderbedarf in Diagnoseförderklassen und ersten Grundschulklassen: Erste Ergebnisse der Mecklenburger Längsschnittstudie. *Zeitschrift für Heilpädagogik*, 8, 282–291.
- Bond, L. (1989). The effects of spatial preparation on measures of scholastic ability. In R. L. Linn (Ed.), *Educational Measurement* (3rd ed., pp. 429–444). New York: American Council on Educational and Macmillan Publishing Co.

- Booth, J.L. & Siegler, R.S. (2006). Developmental and individual differences in pure numerical estimation. *Developmental Psychology*, 42 (1), 189–201.
- Borkowski, J.G. & Peck, V.A. (1986). Causes and consequences of metamemory in gifted children. In R. J. Sternberg & J. E. Davidson (Eds.). *Conceptions of giftedness* (pp. 182–200). Cambridge: Cambridge University Press.
- Bornstein, M.H., Hahn, C.S., Bell, C., Haynes, O.M., Slater, A., Golding, J. & Wolk, D. (2006). Stability in cognition across early childhood. A developmental cascade. *Psychological Science*, 17, 151–158.
- Bortz, J. (2005). *Statistik für Sozialwissenschaftler*. Berlin: Springer.
- Bossaert, G., Doumen, S., Buyse, E. & Verschueren, K. (2011). Predicting children's academic achievement after the transition to first grade. A two-year longitudinal study. *Journal of Applied Developmental Psychology* 32, 47–57.
- Boudon, R. (1974). *Education, opportunity, and social inequality – Changing prospects in Western society*. New York: John Wiley & Sons.
- Brinch, C.N. & Galloway, T.A. (2011). Schooling in adolescence raises IQ scores. *Proceedings of the National Academy of Sciences of the United States of America*, 109 (2), 425–430.
- Brody, N. (1997). Intelligence, schooling, and society. *American Psychologist*, 52, 1046–1050.
- Bronfenbrenner, U. (1986). Ecology of the family as a context for human development: Research perspectives. *Developmental Psychology*, 22, 723–742.
- Brunner, M. (2006). *Mathematische Schülerleistung – Struktur, Schulformunterschiede und Validität* (Dissertation). Verfügbar unter edoc.hu-berlin.de/dissertationen.
- Bull, R. & Johnston, R.S. (1997). Children's arithmetical difficulties: Contributions from processing speed, item identification, and short-term memory. *Journal of Experimental Child Psychology*, 65, 1–24.
- Butterworth, B. (2005). The development of arithmetical abilities. *Journal of Child Psychology and Psychiatry*, 46 (1), 3–18.
- Byrnes, J.P. & Wasik, B.A. Factors predictive of mathematics achievement in kindergarten, first and third grades: An opportunity-propensity analysis. *Contemporary Educational Psychology*, 34, 167–183.
- Cahan, S. & Cohen, N. (1989). Age versus schooling effects on intelligence development. *Child Development*, 60, 1239–1249.
- Cain, K. & Oakhill, J. (2011). Matthew effects in young readers: Reading comprehension and reading experience aid vocabulary development. *Journal of Learning Disabilities*, 44, 431–443.
- Campione, J.C. & Brown, A.L. (1978). Toward a theory of intelligence: Contributions from research with retarded children. *Intelligence*, 2, 279–304.
- Campione, J. C., Brown, A. L. & Ferrara, R. A. (1982). Mental retardation and intelligence. In R. J. Sternberg (Ed.), *Handbook of human intelligence* (pp. 392–490). New York: Cambridge University Press.
- Carroll, J.B. (1963). A model of school learning. *Teachers College Record*, 64, 723–733.

- Carroll, J.B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. New York: Cambridge University Press.
- Carroll, J.B. (1997). The three-stratum theory of cognitive abilities. In D.P. Flanagan, J.L. Genshaft & P.L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests and issues* (pp. 122–130). New York: Guilford Press.
- Cattell, R.B. (1940). A culture free intelligence test, Part I. *Journal of Educational Psychology*, 31, 161–179.
- Cattell, R.B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Educational Psychology*, 54, 1–22.
- Cattell, R.B. (1971). *Abilities: Their structure, growth, and action*. Boston: Houghton Mifflin.
- Cattell, R.B. (1987). *Intelligence: Its structure, growth, and action*. New York: North-Holland.
- Cattell, R.B. & Horn, J. L. (1978). A check on the theory of fluid and crystallized intelligence with description of new subtest designs. *Journal of Educational Measurement*, 15, 139–164.
- Cattell, R.B. & Kline, P. (1977). *The scientific analysis of personality and motivation*. New York: Academic Press.
- Ceci, S.J. (1991). How much does schooling influence general intelligence and its cognitive components? A reassessment of the evidence. *Developmental Psychology*, 27, 703–722.
- Ceci, S.J. & Williams, W.M. (1997). Schooling, intelligence, and income. *American Psychologist*, 52 (10), 1051–1058.
- Colom, R. & Flores-Mendoza, C.E. (2007). Intelligence predicts scholastic achievement irrespective of SES factors: Evidence from Brazil. *Intelligence*, 35, 243–251.
- Cooley, W.W. & Lohnes, P.R. (1976). *Evaluation research in education*. New York: Irvington.
- Cooper, H., Nye, B., Charlton, K., Lindsay, J. & Greathouse, S. (1996). The effects of summer vacation on achievement test scores: A narrative and meta-analytic review. *Review of Educational Research*, 66, 227–268.
- Cragg, L. & Gilmore, C. (2014). Skills underlying mathematics. The role of executive function in the development of mathematics proficiency. *Trends in Neuroscience and Education*, 3, 63–68.
- Crano, W.D., Kenny, D.A. & Campbell, D.T. (1972). Does intelligence cause achievement? A cross-lagged panel analysis. *Journal of Educational Psychology*, 63, 258–275.
- Davis, K., Christodoulou, J., Seidler, S. & Gardner, H. (2011). The Theory of Multiple Intelligences. In R. J. Sternberg & S. B. Kaufman (Eds.), *The Cambridge handbook of intelligence* (pp. 485–503). Cambridge University Press.
- Deary, I.J., Strand, S., Smith, P. & Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence*, 35, 13–21.
- Deary, I.J., Whalley, L. J., Lemmon, H., Crawford, J.R. & Starr, J.M. (2000). The stability of individual differences in mental ability from childhood to old age: Follow-up of the 1932 Scottish Mental Survey. *Intelligence*, 28, 49–55.

- De Corte, E. & Verschaffel, L. (1987). The effect of semantic structure on first graders' solution strategies of elementary addition and subtraction word problems. *Journal for Research in Mathematics Education*, 18, 363–381.
- deGroot, A.D. (1951). War and the intelligence of youth. *Journal of Abnormal and Social Psychology*, 46, 596–597.
- Dehane, S. (1999). *Der Zahlensinn*. Basel: Birkhäuser.
- Devlin, B., Fienberg, S. E., Resnick, D.P. & Roeder, K. (Eds.). (1997). *Intelligence, genes, and success. Scientists respond to the Bell Curve*. New York: Springer-Verlag.
- Ditton, H. & Kreckler, L. (1995). Qualität von Schule und Unterricht. Empirische Befunde zu Fragestellungen und Aufgaben der Forschung. *Zeitschrift für Pädagogik*, 41 (4), 507–529.
- Eid, M., Gollwitzer, M. & Schmitt, M. (2013). *Statistik und Forschungsmethoden. Lehrbuch mit Online-Materialien* (3. korr. Aufl.). Weinheim: Beltz.
- Ericsson, K.A. (1985). Memory skill. *Canadian Journal of Psychology*, 39, 188–231.
- Ericsson, K.A. (1996). *The road to excellence: The acquisition of expert performance in the arts and sciences, sports and games*. New Jersey: Lawrence Erlbaum Associates.
- Ericsson, K.A. & Crutcher, R.J. (1990). The nature of exceptional performance. In P.B. Baltes, D.L. Featherman & R.M. Lerner (Eds.), *Life-span development and behavior* (pp. 187–217). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Ericsson, K.A., Krampe, R.T. & Heizmann, S. (1993). Can we create gifted people? In G. Bock & K. Ackrill (Eds.), *The origin and development of high ability: Ciba Foundation Symposium 178* (pp. 222–249). Chichester: John Wiley & Sons.
- Ericsson, K.A., Krampe, R.T. & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, 100, 363–406.
- Ericsson, K.A. & Smith, J. (1991). *Toward a general theory of expertise: Prospects and limits*. New York: Cambridge University Press.
- Esser, H. (1999). *Soziologie: Spezielle Grundlagen, Band 1: Situationslogik und Handeln*. Frankfurt/M.: Campus.
- Fergusson D.M., Horwood L.J. & Ridder, E.M. (2005). Show me the child at seven: Childhood intelligence and later outcomes in adolescence and young adulthood. *Child Psychology*, 46, 850–858.
- Ferrer, E. & McArdle, J. J. (2004). An experimental analysis of dynamic hypotheses about cognitive abilities and achievement from childhood to early adulthood. *Developmental Psychology*, 40, 935–952.
- Flanagan, D.P. & McGrew, K.S. (1997). A cross-battery approach to assessing and interpreting cognitive abilities: Narrowing the gap between practice and cognitive science. In D. P. Flanagan, J. L. Genshaft & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 314–325). New York: Guilford Press.
- Flanagan, D.P., Andrews, T.J. & Genshaft, J.L. (1997). The functional utility of intelligence tests with special education populations. In D.P. Flanagan, J.L. Genshaft & P.L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 457–483). New York: Guilford Press.

- Flanagan, D.P., McGrew, K.S., & Ortiz, S.O. (2000). *The Wechsler Intelligence Scales and Gf-Gc theory: A contemporary approach to interpretation*. Boston: Allyn & Bacon.
- Floyd, R., G., Evans, J.J. & McGrew, K.S. (2003). Relations between measures of Cattell-Horn-Carroll (CHC) cognitive abilities and mathematics achievement across school-age years. *Psychology in the Schools*, 40, 155–171.
- Floyd, R.G., Reynolds, M.R., Landwirt R.L. & Kranzler, J.H. (2013). Are the general factors from different child and adolescent intelligence tests the same? Results from a five-sample, six-test analysis. *School Psychology Review*, 42, 383–401.
- Flynn, J.R. (1987). Massive IQ gains in 14 nations. What IQ tests really measure. *Psychological Bulletin*, 101 (2), 171–191.
- Franić, S., Dolan, C.V., van Beijsterveldt, C.E.M., Hulshoff Pol, H.E, Bartels, M. & Boomsma, D.I. (2014). Genetic and environmental stability of intelligence in childhood and adolescence. *Twin Research and Human Genetics*, 17 (03), 151–163.
- Frey, M.C. & Dettermann, D.K. (2004). Scholastic assessment or g? The relationship between the Scholastic Assessment Test and general cognitive ability. *Psychological Science*, 15, 373–378.
- Friso-van den Bos, I., van der Ven, S.H.G., Kroesbergen, E.H. & van Luit, J.E.H. (2013). Working memory and mathematics in primary school children. A meta-analysis. *Educational Research Review*, 10, 29–44.
- Fritz, A., Ehlert, A. & Balzer, L. (2013). Development of mathematical concepts as a basis for an elaborated mathematical understanding. *South African Journal of Childhood Education*, 3 (1), 38–67.
- Gagné, F. (2004). Transforming gifts into talents: The DMGT as a developmental theory. *High Ability Studies*, 15, 119–149.
- Gardner, H. (1993). *Frames of mind: The theory of multiple intelligences*. New York: Basic Books.
- Gardner, H. (2006). *Multiple intelligences: New horizons*. New York: Basic Books.
- Gathercole, S.E. & Pickering, S.J. (2000). Working memory deficits in children with low achievements in the national curriculum at 7 years of age. *British Journal of Educational Psychology*, 70, 177–194.
- Gaupp, N., Zoelch, C. & Schumann-Hengsteler, R. (2004). Defizite numerischer Basiskompetenzen bei rechenschwachen Kindern der 3. und 4. Klassenstufe. *Zeitschrift für Pädagogische Psychologie*, 18 (1), 31–42.
- Geary, D.C. (1993). Mathematical disabilities: Cognitive, neuropsychological, and genetic components. *Psychological Bulletin*, 114, 345–362.
- Geary, D.C. (1995). Reflections of evolution and culture in children's cognition. Implications for mathematical development and instruction. *American Psychologist*, 50, 24–37.
- Geary, D.C. (2000). From infancy to adulthood: The development of numerical abilities. *European Child & Adolescent Psychiatry*, 9, II/11– II/16.
- Geary, D.C. (2007). An evolutionary perspective on learning disability in mathematics. *Developmental Neuropsychology*, 32, 471–519.

- Geary, D.C., Brown, S.C. & Samaranayake, V.A. (1991). Cognitive addition: A short longitudinal study of strategy choice and speed-of-processing differences in normal and mathematically disabled children. *Developmental Psychology*, 27, 787–797.
- Geiser, Ch. (2011). *Datenanalyse mit Mplus: Eine anwendungsorientierte Einführung* (2., durchgesehene Auflage). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Goldin, G. & Shteingold, N. (2001). System of representations and the development of mathematical concepts. In A. Cuoco & F. R. Curcio (Eds.), *The roles of representation in school mathematics* (pp. 1-23). Reston, VA: National Council of Teachers of Mathematics.
- Goldstein, H. (2003). *Multilevel statistical models*. London: Arnold.
- Gottfredson, L.S. (1997). Mainstream science on intelligence: An editorial with 52 signatories, history, and bibliography. *Intelligence*, 24 (1), 13–23.
- Gottfredson, L.S. (2001). Intelligence and the American ambivalence toward talent. In N. Colangelo & S. G. Assouline (Eds.), *Talent development IV: Proceedings from the 1998 Henry B. and Jocelyn Wallace National Research Symposium on Talent Development* (pp. 41–58). Scottsdale, AZ: Great Potential Press.
- Gottfredson, L.S. (2003). g, jobs and life. In H. Nyborg (Ed.), *The scientific study of general intelligence: Tribute to Arthur R. Jensen* (pp. 293–342). New York: Pergamon.
- Gustafsson, J.E. (1984). A unifying model for the structure of intellectual abilities. *Intelligence*, 8, 179–203.
- Gustafsson, J.E. (2001). Schooling and intelligence. Effects of track of study on level and profile of cognitive abilities. *International Education Journal*, 2 (4), 166–186.
- Gustafsson, J.E. & Balke, G. (1993). General and specific abilities as predictors of school achievement. *Multivariate Behavioral Research*, 28, 407–434.
- Gustafsson, J.E. & Undheim, J.O. (1992). Stability and change in broad and narrow factors of intelligence from ages 12 to 15 years. *Journal of Educational Psychology*, 84 (2), 141–149.
- Grube, D. (2006a). Entwicklung des Rechnens im Grundschulalter. *Psychologie in Erziehung und Unterricht*, 53, 105-124.
- Grube, D. (2006b). *Entwicklung des Rechnens im Grundschulalter: Basale Fertigkeiten, Wissensabruf und Arbeitsgedächtniseinflüsse* (Reihe Pädagogische Psychologie und Entwicklungspsychologie, Bd. 52). Münster: Waxmann.
- Grube, D. (2008). Rechenschwäche. In W. Schneider & M. Hasselhorn (Hrsg.), *Handbuch der Pädagogischen Psychologie* (Reihe Handbuch der Psychologie, Bd. 10, S. 642–652). Göttingen: Hogrefe.
- Grube, D. & Barth, U. (2004). Rechenleistung bei Grundschulern. Zur Rolle von Arbeitsgedächtnis und basalem Faktenwissen. *Zeitschrift für Pädagogische Psychologie*, 18 (3/4), 245–248.
- Grube, D. & Hasselhorn, M. (2006). Längsschnittliche Analysen zur Lese-, Rechtschreib- und Mathematikleistung im Grundschulalter: Zur Rolle von Vorwissen, Intelligenz, phonologischem Arbeitsgedächtnis und phonologischer Bewusstheit. In I. Hosenfeld & F.-W. Schrader (Hrsg.), *Schulische Leistung: Grundlagen, Bedingungen, Perspektiven* (S. 87–105). Münster: Waxmann.

- Gruber, H. (2008). Lernen und Wissenserwerb. In W. Schneider & M. Hasselhorn (Hrsg.), *Handbuch der Pädagogischen Psychologie* (Reihe Handbuch der Psychologie, Bd. 10, S. 96–98). Göttingen: Hogrefe.
- Gruber, H. & Mandl, H. (1996). Expertise und Erfahrung. In H. Gruber & A. Ziegler (Hrsg.), *Expertiseforschung. Theoretische und methodische Grundlagen* (S. 18–34). Opladen: Westdeutscher Verlag.
- Hany, E.A. & Heller, K.A. (1991). Gegenwärtiger Stand der Hochbegabungsforschung. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 23, 241–249.
- Härnqvist, K. (1968). Relative changes in intelligence from 13–18: Results. *Scandinavian Journal of Psychology*, 9, 50–64.
- Haertel, G.D., Walberg, H.J., & Weinstein, T. (1983). Psychological models of educational performance: A theoretical synthesis of constructs. *Review of Educational Research*, 53, 75–91.
- Hambrick, D.Z. (2004). The role of domain knowledge in higher-level cognition. In O. Wilhelm, & R.W. Engle (Eds.), *Handbook of understanding and measuring intelligence* (pp. 361–372). Thousand Oaks, CA: Sage Publications.
- Hambrick, D.Z. & Engle, R.W. (2002). Effects of domain knowledge, working memory capacity, and age on cognitive performance: An investigation of the knowledge-is-power hypothesis. *Cognitive Psychology*, 44, 339–387.
- Hartke, B., Koch, K. & Blumenthal, Y. (2010). *Wie effektiv sind Diagnoseförderklassen? Zur Wirksamkeit des Unterrichts mit schulisch gefährdeten Kindern in Grundschulklassen und in Diagnoseförderklassen (DFK) – Ergebnisse der Mecklenburger Längsschnittstudie*. Bericht zu den Ergebnissen der wissenschaftlichen Begleitstudie des Projekts „Primarstufe“ – Teil II. Universität Rostock.
- Hasselhorn, M. & Grube, D. (1997). Entwicklung der Intelligenz und des Denkens. In F.W. Weinert & Hekmke (Hrsg.), *Entwicklung im Grundschulalter* (S. 15–26). Weinheim: Beltz PVU.
- Hasselhorn, M. & Grube, D. (2006). Gedächtnisentwicklung (Grundlagen). In W. Schneider & B. Sodian (Hrsg.), *Enzyklopädie der Psychologie. Themenbereich C: Theorie und Forschung, Serie V: Entwicklung, Band 2: Kognitive Entwicklung* (S. 271–325). Göttingen: Hogrefe.
- Hattie, J. (2012). *Visible learning: A synthesis of over 800 meta-analyses relating to student achievement*. London: Routledge.
- Heller, K.A. (1973). *Intelligenzmessung*. Villingen: Neckarverlag.
- Heller, K.A. (Hrsg.) (1984). *Leistungsdiagnostik in der Schule* (4. Aufl.). Bern: Hans Huber.
- Heller, K.A. (1991). Schuleignungsprognostik. In K.A. Heller (Hrsg.), *Begabungsdiagnostik in der Schul- und Erziehungsberatung* (4. Aufl., S. 213–235). Bern: Huber.
- Heller, K.A. (Hrsg.) (1992). *Hochbegabung im Kindes- und Jugendalter*. Göttingen: Hogrefe.
- Heller, K.A. (1997). Individuelle Bedingungsfaktoren der Schulleistung. In Weinert, F.E. & Helmke, A. (Hrsg.), *Entwicklung im Grundschulalter* (S. 181–203) Weinheim: Beltz.
- Heller, K.A. (Hrsg.) (2001). *Hochbegabung im Kindes- und Jugendalter* (2., überarbeitete und erweiterte Aufl.). Göttingen: Hogrefe.

- Heller, K.A. (2008). *Von der Aktivierung der Begabungsreserven zur Hochbegabtenförderung*. Berlin: LIT.
- Heller, K.A., Schön-Gaedike, A.-K. & Weinländer, H. (1976). *Kognitiver Fähigkeits-Test für 4. Bis 13. Klassen (KFT 4-13+)*. Weinheim: Beltz.
- Heller, K.A. & Perleth, Ch. (2000). *Kognitiver Fähigkeitstest für 4. bis 12. Klassen, Revision (KFT 4-12+ R)*. Göttingen: Beltz.
- Helmke, A. (1997). Individuelle Bedingungsfaktoren der Schulleistung: Ergebnisse aus dem SCHOLASTIK-Projekt. In F.E. Weinert & A. Helmke (Hrsg.), *Entwicklung im Grundschulalter* (S. 203-215). Weinheim: Beltz.
- Helmke, A., Rindermann, H. & Schrader, F.W. (2008). Wirkfaktoren akademischer Leistungen in Schule und Hochschule. In W. Schneider & M. Hasselhorn (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (S. 145–155). Göttingen: Hogrefe.
- Helmke, A. & Schrader, F.-W. (2010). Determinanten der Schulleistung. In D. H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (S. 90–102). Weinheim: Beltz.
- Helmke, A. & Weinert, F.E. (1997). Bedingungsfaktoren schulischer Leistungen. In F.E. Weinert (Hrsg.), *Enzyklopädie der Psychologie, Serie Pädagogische Psychologie, Psychologie des Unterrichts und der Schule* (Bd. 3, 71–176). Göttingen: Hogrefe.
- Herrnstein, R.J. & Murray, Ch. (1994). *The bell curve: Intelligence and class structure in American life*. New York: Free Press.
- Hesse, I. & Latzko, B. (2011). *Diagnostik für Lehrkräfte* (2. Aufl.). Opladen: Budrich UTB.
- Hildebrandt, A., Wilhelm, O. & Robitzsch, A. (2009). Complementary and competing factor analytic approaches for the investigation of measurement invariance. *Review of Psychology*, 16 (2), 87–102.
- Holling, H., Preckel, F. & Vock, M. (2004). *Intelligenzdiagnostik - Kompendien Psychologische Diagnostik, Bd. 6*. Göttingen: Hogrefe.
- Holmes, J. & Adams, J.W. (2006). Working memory and children's mathematical skills: Implications for mathematical development and mathematics curricula. *Educational Psychology*, 26, 339–366.
- Hofstaetter P.R. (1954). The changing composition of "intelligence": A study in t-technique. *The Journal of Genetic Psychology*, 85 (1), 159–164.
- Horn, J. L. (1994). The theory of fluid and crystallized intelligence. In R.J. Sternberg (Ed.), *Encyclopedia of human intelligence* (pp. 433–451). New York, NY: McMillan.
- Horn, J. L. & Cattell, R.B. (1966). Refinement and test of the theory of fluid and crystallized intelligence. *Journal of Educational Psychology*, 57, 253–270.
- Horn, J.L. & Cattell, R.B. (1967). Age differences in fluid and crystallized intelligence. *Acta Psychologica*, 26 (2), 107–129.
- Horn, J.L. & Noll, J. (1997). Human cognitive capabilities: Gf–Gc theory. In D. P. Flanagan, J. L. Genshaft & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests and issues* (pp. 53–91). New York: Guilford Press.
- Husén, T. & Tuijnman, A. (1991). The contribution of formal schooling to the increase in intellectual capital. *Educational Researcher*, 20 (1), 17–25.

- Ingenkamp, K. & Lissmann, U. (2008). *Lehrbuch der pädagogischen Diagnostik* (6. Aufl.). Weinheim: Beltz.
- Jäger, A.O., Süß, H.M. & Beauducel, A. (1996). *Test für das Berliner Intelligenzstrukturmodell (BIS)*. Göttingen: Hogrefe.
- Jäger, A.O., Süß, H. M. & Beauducel, A. (1997). *Berliner Intelligenzstruktur-Test (BIS-Test), Form 4. Handanweisung*. Göttingen: Hogrefe.
- Jäger, R.S. (2007). *Beobachten, beurteilen, fördern. Lehrbuch für die Aus-, Fort- und Weiterbildung*. Landau: Empirische Pädagogik.
- Jencks, C., Smith, M., Acland, H., Bane, M. J., Cohen D., Gintis, H., Heyns, B. & Mitchelson, S. (1972). *Inequality: A reassessment of the effects of family and schooling in America*. New York: Basic Books.
- Jensen, A.R. (1998). *The g factor: The science of mental ability*. New York: Praeger.
- Jensen, A.R. (2000, August). The g factor and the design of education. *Paper presented at the annual meeting of the American Psychological Association*, Washington DC.
- Jöreskog, K.G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, 34, 183–202.
- Jöreskog, K.G. (1979a). A general approach to confirmatory maximum likelihood factor analysis, with addendum. In K.G. Jöreskog & D. Sörbom, *Advances in factor analysis and structural equation models* (J. Magidson, Ed., pp. 21–43). Cambridge, MA: Abt Books.
- Jöreskog, K.G. (1979b). Structural equation models in the social sciences: Specification, estimation and testing. In K.G. Jöreskog & D. Sörbom, *Advances in factor analysis and structural equation models* (J. Magidson, Ed., pp. 105–127). Cambridge, MA: Abt Books.
- Johnson, W. & Bouchard Jr., T.J. (2005). The structure of human intelligence: It is verbal, perceptual and image rotation (VPR), not fluid and crystallized. *Intelligence*, 33, 431–444.
- Johnson, W., Bouchard Jr., T.J., Krueger, R.F., McGue, M. & Gottesman, I.I. (2004). Just one g: Consistent results from three test batteries. *Intelligence*, 32, 95–107.
- Kamphaus, R. W. (2001). *Clinical assessment of child and adolescent intelligence* (2nd ed.). Boston: Allyn and Bacon.
- Kan, K.J., Kievit, R.A., Dolan, C., van der Maas, H. (2011). On the interpretation of the CHC factor Gc. *Intelligence*, 39, 292–302.
- Karbach, J., Gottschling, J., Spengler, M., Hegewald, K. & Spinath, F.M. (2013). Parental involvement and general cognitive ability as predictors of domain-specific academic achievement in early adolescence. *Learning and Instruction*, 23, 43–51.
- Kaufman, S.B., DeYoung, C.G., Gray J.R., Brown, J. & Mackintosh, N. (2009). Associative learning predicts intelligence above and beyond working memory and processing speed. *Intelligence*, 37, 374–382.
- Kaufman, S.B., Reynolds M.R., Liu, X. L., Kaufman, A.S. & McGrew, K.S. (2012). Are cognitive g and academic achievement g one and the same g? An exploration on the Woodcock–Johnson and Kaufman tests. *Intelligence*, 40 (2), 123–138.

- Kaya, F., Juntune, J. & Stough, L. (2015). Intelligence and its relationship to achievement. *Elementary Education Online*, 14 (3), 1060–1078.
- Kline, R.B. (1998) Principles and practice of structural equation modeling. In Ch. Geiser (2011). *Datenanalyse mit Mplus*. 2., durchgesehene Auflage. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Klieme, E. & Rakoczy, K. (2008). Empirische Unterrichtsforschung und Fachdidaktik: Outcome-orientierte Messung und Prozessqualität des Unterrichts. *Zeitschrift für Pädagogik*, 54, (2), 222–237.
- Klieme, E. Schümer, G. & Knoll, S. (2001). Mathematikunterricht in der Sekundarstufe I: „Aufgabekultur“ und Unterrichtsgestaltung im internationalen Vergleich. In E. Klieme E. & J. Baumert (Hrsg), *TIMSS – Impulse für Schule und Unterricht* (S. 43–57). Bonn: Bundesministerium für Bildung und Forschung.
- Knight, G.P. & Zerr, A. A. (2010). Informed theory and measurement equivalence in child development research. *Child Development Perspectives*, 4 (1), 25–30.
- Koch, K., Hartke, B. & Blumenthal, Y. (2008). *Die Lernausgangslage von Kindern mit besonderem Förderbedarf in Grundschulklassen 1 und Diagnoseförderklassen. Erste Ergebnisse der wissenschaftlichen Begleitstufe des Projekts "Primarstufe"* (Forschungsbericht). Universität Rostock.
- Koch, K., Hartke, B. & Blumenthal, Y. (2009). *Merkmale von Kindern mit besonderem Förderbedarf im ersten Schuljahr – Erste Ergebnisse der Mecklenburger Längsschnittstudie*. Sonderpädagogik in Forschung und Praxis. Bd. 26. Dr. Kovac: Hamburg.
- Koch, K. Hartke, B. & Blumenthal, Y. (2010). *Wie effektiv sind Diagnoseförderklassen? Zur Wirksamkeit des Unterrichts mit schulisch gefährdeten Kindern in Grundschulklassen und in Diagnoseförderklassen (DFK). Ergebnisse der Mecklenburger Längsschnittstudie. Bericht zu den Ergebnissen der wissenschaftlichen Begleitstudie des Projekts „Primarstufe“ – Teil II* (Forschungsbericht). Universität Rostock. Verfügbar unter http://www.sopaed.uni-rostock.de/fileadmin/Isoheilp/2_dfk_bericht.pdf
- Koening, K.A., Frey, M. C. & Dettermann, D.K. (2008). ACT and general cognitive ability. *Intelligence*, 36, 153–160.
- Köller, O. (2008). Forschungsansätze in der pädagogischen Psychologie. In W. Schneider & M. Hasselhorn (Hrsg.), *Handbuch der pädagogischen Psychologie* (S. 297–712). Göttingen: Hogrefe.
- Köller, O. & Baumert, J. (2002). Entwicklung schulischer Leistungen. In R. Oerter & L. Montada (Hrsg.), *Entwicklungspsychologie* (5., vollst. neu bearb. Aufl., S. 756–786). Weinheim: Beltz.
- Koriakin, T., White, E., Breaux, K. C., DeBiase, E., O'Brien, R., Howell, M., Costa, M., Liu, X., Pan, X. & Courville, T. (2016). Patterns of cognitive strengths and weaknesses and relationships to math errors. *Journal of Psychoeducational Assessment*, 35 (1-2), 155–167.
- Krajewski, K. (2003). *Vorhersage von Rechenschwäche in der Grundschule*. Hamburg: Kovac.
- Krajewski, K. (2008a). Prävention der Rechenschwäche. In W. Schneider & M. Hasselhorn (Hrsg.), *Handbuch der Psychologie, Bd. Pädagogische Psychologie* (S. 360–370). Göttingen: Hogrefe.

- Krajewski, K. (2008b). Vorschulische Förderung mathematischer Kompetenzen. In F. Pe-
termann & W. Schneider (Hrsg.), *Enzyklopädie der Psychologie, Reihe Entwicklungs-
psychologie, Bd. Angewandte Entwicklungspsychologie* (S. 275–304). Göttingen: Ho-
greffe.
- Krajewski, K. (2008). Mathematische Vorläuferfertigkeiten im Vorschulalter. In W. Schnei-
der. *Handbuch der Pädagogischen Psychologie*. Göttingen: Hogrefe.
- Krajewski, K., Küspert, P. & Schneider, W. (2002). *Deutscher Mathematiktest für erste
Klassen (DEMAT 1+)*. Göttingen: Hogrefe.
- Krajewski, K., Liehm, S. & Schneider, W. (2004). *Deutscher Mathematiktest für zweite
Klassen (DEMAT 2+)*. Göttingen: Hogrefe.
- Krajewski, K. & Schneider, W. (2006). Mathematische Vorläuferfertigkeiten im Vorschul-
alter und ihre Vorhersagekraft für die Mathematikleistungen bis zum Ende der Grund-
schulzeit. *Psychologie in Erziehung und Unterricht*, 53, 246–262.
- Krapp, A. & Weidenmann, B. (2001). *Pädagogische Psychologie* (4., vollst. überarbeitete
Aufl.) Weinheim: Beltz.
- Krumm, S., Ziegler, M. & Bühler, M. (2008). Reasoning and working memory as predictors
of school grades. *Learning and Individual Differences*, 18 (2), 248–257.
- Kvist, A. V. & Gustafsson, J. E. (2008). The relation between fluid intelligence and the
general factor as a function of cultural background: A test of Cattell's investment theory.
Intelligence, 36 (5), 422–436.
- Laidra, K., Pullmann, H. & Allik, J. (2007). Personality and intelligence as predictors of
academic achievement: A cross-sectional study from elementary to secondary school.
Personality and Individual Differences, 42, 441–451.
- Langfeldt, H.-P. (2006). *Psychologie für die Schule* (1. Aufl.). Weinheim, Basel: Beltz.
- Lehnhard, W. & Lehnhard, A. (2010). *Rechenspiele mit Elfe und Mathis I: Ein Mathema-
tiktraining der ersten bis dritten Jahrgangsstufe*. Hogrefe: Göttingen.
- Lemos, G.C., Almeida, L.S. & Colom, R. (2011). Intelligence of adolescents is related to
their parents' educational level but not to family income. *Personality and Individual Dif-
ferences*, 50 (7), 1062–1067.
- Lienert, G.A. & Raatz, U. (1994). *Testaufbau und Testanalyse* (5. Aufl.). Weinheim: Beltz.
- Lohman, D.F. (1999). Minding our p's and q's: On finding relationships between learning
and intelligence. In P.L. Ackerman, P.C. Kyllonen & R. D. Roberts (Eds.), *Learning and
individual differences: Process, trait, and content determinants* (pp. 55–76). Washing-
ton, DC: American Psychological Association.
- Lorenz, J.H. (2009). Diagnose und Prävention von Rechenschwäche als Herausforderung
im Elementar- und Primarbereich. In A. Heinze & M. Grüßing (Hrsg.), *Mathematikler-
nen vom Kindergarten bis zum Studium* (S. 17–34). Münster: Waxmann.
- Lubinski, D. & Dawis, R.V. (1992). Aptitudes, skills, and proficiencies. In M.D. Dunnette
& L.M. Hough (Eds.), *The handbook of industrial/organizational psychology* (2nd ed.,
pp. 1–59). Palo Alto: Consulting Psychologists Press.
- Lüdtke, O. & Robitzsch, A. (2010). *Umgang mit fehlenden Daten in der empirischen Bil-
dungsforschung*. Weinheim: Juventa.

- Lukesch, H. (1998). *Einführung in die pädagogisch-psychologische Diagnostik*. Regensburg: Roderer.
- Luo, D., Thompson, L.A. & Detterman, D.K. (2006a). The causal factor underlying the correlation between psychometric g and scholastic performance. *Intelligence*, 31, 67–83.
- Luo, D., Thompson, L.A. & Detterman, D.K. (2006b). The criterion validity of tasks of basic cognitive processes. *Intelligence*, 34, 79–120.
- Lynn, R. & Vanhanen, T. (2012). *Intelligence: A unifying construct for social sciences*. London: Ulster Institute for Social Research.
- Mackintosh, N.J. (2011). *IQ and human intelligence*. Oxford, UK: Oxford University Press.
- Martin, R.B., Ewing-Cobbs, L. & Fletcher, J.M. (2012) Prediction and stability of mathematics skill and difficulty. *Journal of Learning Disabilities*, 46 (5) 428–443.
- Mayer, R.E. (1992). *Thinking, problem solving, cognition*. New York: W.H. Freeman and Company.
- Merz, F., Remer, H. & Ehlers, T. (1985). Der Einfluß des Schulbesuchs auf Intelligenztestleistungen im Grundschulalter. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 17 (3), 223–241.
- Messick, S. (1984). The psychology of educational measurement. *Journal of Educational Measurement*, 21, 215–237.
- McArdle, J.J. (1994). Structural factor analysis experiments with incomplete data. *Multivariate Behavioral Research*, 29, 409–454.
- McArdle, J.J. (2009). Latent variable modeling of differences and changes with longitudinal data. *Annual Review of Psychology*, 60, 577–605.
- McArdle, J.J. & Hamagami, F. (2001). Latent difference score structural models for linear dynamic analyses with incomplete longitudinal data. In L.M. Collins & A.G. Sayer (Eds.), *New methods for the analysis of change* (pp. 139–175). Washington, DC: American Psychological Association.
- McArdle, J. J. & Woodcock, R.W. (1997). Expanding test-retest designs to include developmental time-lag components. *Psychological Methods*, 2, 403–435.
- McCall, G.J. (1984). Systematic Field Observation. *Annual Review of Sociology*, 10, 263–282.
- McGrew, K.S. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence*, 37, 1–10.
- McGrew, K., Flanagan, D.P., Keith, T.Z. & Vanderwood, M. (1997). Beyond g: The Impact of Gf-Gc specific cognitive abilities research on the future use and interpretation of intelligence tests in the Schools. *School Psychology Review*, 26 (2), 189–210.
- Ministerium für Bildung, Wissenschaft und Kultur Mecklenburg-Vorpommern (2010). Schulgesetz für das Land Mecklenburg-Vorpommern (Schulgesetz - SchulG M-V) in der Fassung der Bekanntmachung vom 10. September 2010. Stand: letzte berücksichtigte Änderung: §§ 45, 46 geändert durch Gesetz vom 17. Dezember 2015 (GVBl. M-V S. 586).
- Muthén, L.K. & Muthén, B.O. (2012). *MPlus user's guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.

- Naglieri, J.A. & Bornstein, B. (2003). Intelligence and achievement: Just how correlated are they? *Journal of Psychoeducational Assessment*, 21, 244–260.
- Neisser, U., Boodoo, G., Bouchard, T.J., Boykin, A.W., Brody, N., Ceci, S.J., Halpern, D.F., Loehlin, J.C., Perloff, R., Sternberg, R.J., Urbina, S. (1996). Intelligence: Knowns and unknowns. *American Psychologist*, 51 (2), 77–101.
- Otto, S. (2012). *Zur Entwicklung und Stabilität von Intelligenz im frühen Grundschulalter: Eine empirische Untersuchung mit dem KFT 1-2* (Dissertation). Universität Rostock.
- OECD (2002). *PISA 2000 technical report*. Paris: OECD.
- OECD (2003). *PISA 2003 assessment framework: Mathematics, reading, science and problem solving knowledge and skills*. Paris: OECD.
- OECD (2004a). *Learning for tomorrow's world. First results from PISA 2003*. Paris: OECD.
- OECD (2004b). *Problem solving for tomorrow's world*. Paris: OECD.
- Passolunghi, M.C. & Siegel, L.S. (2001). Short-term memory, working memory, and inhibitory control in children with difficulties in arithmetic problem solving. *Journal of Experimental Child Psychology*, 80, 44–57.
- Perels, F., Otto, B. & Schmitz, B. (2008). Spezielle Auswertungsmethoden der pädagogischen Psychologie. In W. Schneider & M. Hasselhorn (Hrsg.), *Handbuch der pädagogischen Psychologie* (S. 712–720). Göttingen: Hogrefe.
- Perleth, Ch. (1997). *Zur Rolle von Begabung und Erfahrung bei der Leistungsgenese. Ein Brückenschlag zwischen Begabungs- und Expertiseforschung* (unveröffentl. Habilitationsschrift). München: LMU.
- Perleth, Ch. (2001). Follow-up-Untersuchungen zur Münchner Hochbegabungsstudie. In K.A. Heller (Hrsg.), *Hochbegabung im Kindes- und Jugendalter* (2. Aufl., S. 357–446) Göttingen: Hogrefe.
- Perleth, Ch. & Heller, K.A. (1994). The Munich longitudinal study of giftedness. In R.F. Subotnik & K.D. Arnold (Eds.), *Beyond Terman: Contemporary longitudinal studies of giftedness and talents* (pp. 77–114). Norwood, NJ: Ablex.
- Perleth, Ch. & Heller, K.A. (2008). *Kognitiver Fähigkeitstest für die 3. Klasse. (KFT3 R)* (Unveröffentlichte Revision). Universität Rostock.
- Perleth, Ch. & Heller, K.A. (in Vorb.). *Kognitiver Fähigkeitstest für Vorschulalter bis 3. Klasse – Revision* (KFT K-3 R). Göttingen: Beltz.
- Perleth, Ch., Joswig, H. & Hoese, D. (2012). Beratungspraxis der Begabungspsychologischen Beratungsstelle des Odysseus-Projekts am Institut für Pädagogische Psychologie „Rosa und David Katz“ der Universität Rostock. In A. Ziegler, R. Grassinger & B. Harder (Hrsg.), *Konzepte der Hochbegabtenberatung in der Praxis* (S. 129–156). Berlin: LIT-Verlag.
- Perleth, Ch. & Sierwald, W. (2001). Zur Entwicklung von Intelligenz und Kreativität bei hochbegabten Kindern und Jugendlichen. In K. A. Heller (Hrsg.), *Hochbegabung im Kindes- und Jugendalter* (2. überarb. und erweit. Aufl., S. 171–355). Göttingen: Hogrefe..

- Perleth, Ch. & Sierwald, W. (1996). *Stabilität und Veränderung in einem zweidimensionalen Intelligenzstrukturmodell für den KFT 4-13. Ein Beitrag zur psychometrischen Intelligenzforschung* (Unveröffent. Forschungsbericht). München: LMU.
- Piaget, J. (1969). *Das Erwachen der Intelligenz beim Kinde*. Stuttgart: Klett.
- Piaget, J. (1972). Intellectual development from adolescence to adulthood. *Human Development*, 15, 1–15.
- Piaget, J. (1976). *Die Äquilibration der kognitiven Strukturen*. Stuttgart: Klett.
- PISA Programme for International Student Assessment (Programm zur Internationalen Bewertung von Schülerleistungen) Laufende OECD-Studie (1998--2007) zur Lesekompetenz, zur mathematisch/naturwissenschaftlichen Grundbildung und zu fächerübergreifenden Kompetenzen mit vielfältigen Indikatoren für Lernergebnisse und ihre Bedingungen bei 15jährigen Schülern. MPIB, Berlin.
- PISA - Schülerleistungen im internationalen Vergleich. (2000). Eine neue Rahmenkonzeption für die Erfassung von Wissen und Fähigkeiten. Deutsches PISA-Konsortium (Hrsg.). OECD Programme for International Student Assessment. Max-Planck-Institut für Bildungsforschung: Berlin.
- Prenzel, M., Baumert, J., Blum, W., Lehmann, R., Leutner, D., Neubrand, M., Pekrun, R., Rolff, H.-G., Rost, J. & Schiefele, U. (Hrsg.) (2004). *PISA 2003. Ergebnisse des zweiten internationalen Vergleichs*. Münster: Waxmann.
- Primi, R. (2002). Complexity of geometric inductive reasoning tasks: Contribution to the understanding of fluid intelligence. *Intelligence*, 30, 41–70.
- Primi, R., Ferrão, M.E. & Almeida, L.S. (2010). Fluid intelligence as a predictor of learning: A longitudinal multilevel approach applied to math. *Learning and Individual Differences*, 20 (5), 446–451.
- Raven, J.C. (1938). *Standard Progresssive Matrices. A perceptual test of intelligence*. London: H.K. Lewis.
- Raven, J.C. (2009). *Standard Progresssive Matrices. A. perceptual test of intelligence*. Parson.
- Renkl, A. (1996). Vorwissen und Schulleistung. In J. Möller & O. Köller (Hrsg.), *Emotionen, Kognitionen und Schulleistung* (S. 175–190). Weinheim: Beltz.
- Renkl, A. & Stern, E. (1994). Die Bedeutung von kognitiven Eingangsvoraussetzungen und Lernaufgaben für das Lösen von einfachen und komplexen Textaufgaben. *Zeitschrift für Pädagogische Psychologie*, 8, (1), 27–39.
- Reynolds, M.R., Keith, T.Z., Flanagan, D.P. & Alfonso, V.C. (2013). A cross-battery, reference variable, confirmatory factor analytic investigation of the CHC taxonomy. *Journal of School Psychology*, 51 (4), 535–555.
- Rindermann, H. (2006). Was messen internationale Schulleistungsstudien? *Psychologische Rundschau*, 57 (2), 69–86.
- Rindermann, H., Flores-Medozza, C. & Mansur-Alves, M. (2010). Reciprocal effects between fluid and crystallized intelligence and their dependence on parents' socioeconomic status and education. *Learning and Individual Differences*, 20, 544–548.

- Rindermann, H. & Neubauer, A.C. (2004). Processing speed, intelligence, creativity, and school performance: Testing of causal hypotheses using structural equation models. *Intelligence*, 32(6), 573–589.
- Rohde, T.E. & Thompson, L.A. (2007). Predicting academic achievement with cognitive ability. *Intelligence*, 35, 83–92.
- Roick, T., Gölit, D. & Hasselhorn, M. (2004). *Deutscher Mathematiktest für dritte Klassen (DEMAT 3+)*. Göttingen: Beltz.
- Roick, T., Gölit, D. & Hasselhorn, M. (2006). *Deutscher Mathematiktest für vierte Klassen (DEMAT 4)*. Göttingen: Beltz.
- Rönnlund, M., Sundström, A. & Nilsson, L.-G. (2015). Interindividual differences in general cognitive ability from age 18 to age 65 years are extremely stable and strongly associated with working memory capacity. *Intelligence*, 53, 59–64.
- Rost, D.H. (2005). *Interpretation und Bewertung pädagogisch-psychologischer Studien*. Weinheim: Beltz UTB.
- Rost, D.H. (2010). Stabilität von Hochbegabung. In F. Preckel, W. Schneider & H. Holling (Hrsg.), *Diagnostik von Hochbegabung* (S. 239–266). Göttingen: Hogrefe.
- Sanchez, M.D.P. & Sternberg, R.J. (1991). La teoria triaica de la inteligencia: Un modelo que ayuda entender la naturaleza del retraso mental. *Revista interuniversitaria de formación del profesorado*, 11, 77–93.
- Sass, D.A. & Schmitt, T.H. (2013). Testing measurement and structural invariance: Implications for practice. In T. Teo (Ed.), *Handbook of quantitative methods for educational research* (pp. 315–345). Auckland: Springer.
- Sauer, J. (2001). Prognose des Schulerfolgs. In D.H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (2. Aufl., S. 584–595) Weinheim: Beltz.
- Sauer, J. & Gamsjäger, E. (1996). *Ist Schulerfolg vorhersehbar? Die Determinanten der Grundschulleistung und ihr prognostischer Wert für den Sekundarschulerfolg*. Göttingen: Hogrefe.
- Schäfer, B.A. & McDermott, P.A. (1999). Learning behavior and intelligence as explanations for children's scholastic achievement. *Journal of School Psychology*, 37 (3), 299–313.
- Schipolowski, S., Wilhelm, O. & Schroeders, U. (2014). On the nature of crystallized intelligence: The relationship between verbal ability and factual knowledge. *Intelligence*, 46, 156–168.
- Schmiedek, F. & Li, S.C. (2015). Intelligence: Central conceptions and psychometric models. In J. D. Wright (Ed.), *International Encyclopedia of the Social & Behavioral Sciences* (2nd edition, pp. 290–296). Amsterdam: Elsevier.
- Schneider, W. (1992). Erwerb von Expertise: Zur Relevanz kognitiver und nicht-kognitiver Voraussetzungen. In E.A. Hany & H. Nickel (Hrsg.), *Begabung und Hochbegabung: Theoretische Konzepte – empirische Befunde – praktische Konsequenzen* (S. 105–122). Bern: Verlag Hans Huber.
- Schneider, W. (1993). Acquiring expertise: Determinants of exceptional performance. In K. A. Heller, F. J. Mönks & A. H. Passow (Hrsg.), *International handbook of research and development of giftedness and talent* (pp. 311–324). Oxford: Pergamon Press.

- Schneider, W. (2008). Entwicklung der Intelligenz und des Denkvermögens in Kindheit, Jugend und Erwachsenenalter. In W. Schneider (Hrsg.), *Entwicklung von der Kindheit bis zum Erwachsenenalter. Befunde der Münchner Längsschnittstudie LOGIK* (S. 43–66). Weinheim: Beltz PVU.
- Schneider, W. (2013). Principles of assessment of aptitude and achievement. In D. H. Saklofske, V. L. Schwane & C. R. Reynolds (Eds.), *The Oxford handbook of child psychological assessment* (pp. 286–330). New York: Oxford.
- Schneider, W., Bullock, M. & Sodian, B. (1998). Die Entwicklung des Denkens und der Intelligenzunterschiede zwischen Kindern. In F. E. Weinert (Hrsg.), *Entwicklung im Kindesalter* (S. 53–74). Weinheim: Beltz PVU.
- Schneider, W. & Hasselhorn, M. (Hrsg.). (2007). *Handbuch der Entwicklungspsychologie. Entwicklung der Intelligenz im Kindesalter*. Göttingen: Hogrefe.
- Schneider, W., Küspert, P. & Krajewski, K. (2013). *Die Entwicklung mathematischer Kompetenzen*. Paderborn: UTB.
- Schneider, W. & Lindenberger, U. (Hrsg.). (2012). *Entwicklungspsychologie*. Weinheim: Beltz.
- Schneider, W.J. & McGrew, K.S. (2012). The Cattell–Horn–Carroll model of intelligence. In D. Flanagan & P. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 99–144). New York: Guilford Press.
- Schneider, W., Niklas, F. & Schmiedeler, S. (2014). Intellectual development from early childhood to early adulthood. The impact of early IQ differences on stability and change over time. *Learning and Individual Differences*, 32, 156–162.
- Schneider, W. & Stefanek, J. (2004). Entwicklungsveränderungen allgemeiner kognitiver Fähigkeiten und schulbezogener Fertigkeiten im Kindes- und Jugendalter. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 36 (3), 147–159.
- Schroeders, U., Schipolowski, S., Zettler, I., Golle, J. & Wilhelm, O. (2016). Do the smart get smarter. Development of fluid and crystallized intelligence in 3rd grade. *Intelligence*, 59, 84–95.
- Sekretariat der ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (2004).
- Shanley, L. (2016). Evaluating longitudinal mathematics achievement growth: Modeling and measurement considerations for assessing academic progress. *Educational Researcher*, 45 (6), 347–357.
- Silver, C.H., Pennett, H.D., Black, J.L., Fair, G.W. & Balise, R.R. (1999). Stability of arithmetic disability subtypes. *Journal of Learning Disabilities*, 2, 108–119.
- Snow, R.E. (1989). Aptitude, instruction, and individual development. *International Journal of Educational Research*, 13, 869–881.
- Snyderman, M. & Rothman, S. (1987). Survey of expert opinion on intelligence and aptitude testing. *American Psychologist*, 42, 137–144.
- Soares, D.L., Lemos, G.C., Primi, R. & Almeida, L.S. (2015). The relationship between intelligence and academic achievement throughout middle school: The role of students' prior academic performance. *Learning and Individual Differences*, 41, 73–78.

- Spada, H. & Wichmann, S. (1996). Kognitive Determinanten der Lernleistung. In F.E. Weinert (Hrsg.), *Psychologie des Lernens und der Instruktion* (S. 119–152). Göttingen: Hogrefe.
- Spearman, C. (1904). "General intelligence," objectively determined and measured. *American Journal of Psychology*, 15, 1–199.
- Spearman, C. (1927). *The abilities of man, their nature and measurement*. New York: Macmillan.
- Spelke, E.S. (2000). Core Knowledge. *American Psychologist*, 55 (11), 1233–1243.
- Spinath B. & Spinath F.M., Harlaar, N. & Plomin, R. (2006). Predicting school achievement from general cognitive ability, self-perceived ability, and intrinsic value. *Intelligence*, 34, 363–374.
- Stanovich, K.E. (1986). Matthew effects in reading. Some consequences of individual differences in the acquisition of literacy. *Research Quarterly*, 21 (4), 360–407.
- Stanovich, K.E. (1986). *How to think straight about psychology*. Glenview, IL: Scott, Foresman and Company.
- Stern, E. (1997). Erwerb mathematischer Kompetenzen: Ergebnisse. In F. E. Weinert & A. Helmke (Hrsg.), *Entwicklung im Grundschulalter* (S. 159–162). Weinheim: Beltz.
- Stern, E. (2004). Lernen ist der mächtigste Mechanismus der kognitiven Entwicklung. *Jahrbuch 2003/2004 der Max-Planck-Gesellschaft*, 1–6.
- Stern, E. (2008). Verpasste Chancen? Was wir aus der LOGIK-Studie über den Mathematikunterricht lernen können. In W. Schneider (Hrsg.), *Entwicklung von der Kindheit bis zum Erwachsenenalter* (S. 187–202). Weinheim: Beltz.
- Stern, W. (1912). *Die psychologischen Methoden der Intelligenzprüfung und deren Anwendung an Schulkindern*. Leipzig: Ambrosius.
- Sternberg, R.J. (1985). *Beyond IQ: A triarchic theory of human intelligence*. Cambridge: Cambridge University Press.
- Sternberg, R.J. (1986). A triarchic theory of intellectual giftedness. In R. J. Sternberg & J. E. Davidson (Eds.), *Conceptions of giftedness* (pp. 223–243). Cambridge: Cambridge University Press.
- Sternberg, R.J. & Detterman, D.K. (1986). *What is intelligence? Contemporary viewpoints on its nature and definition*. Michigan: Ablex.
- Sternberg, R.J. & Powell, J.S. (1983). The development of intelligence. In J. H. Flavell & E. Markman (Eds.), *Handbook of child psychology: Vol. III. Cognitive development* (4th ed., pp. 341–419). New York: Wiley.
- Sternberg, R.J. & Kaufman, S.B. (Eds.) (2011). *The Cambridge handbook of intelligence*. Cambridge University Press.
- Stumpf, E. & Schneider, W. (2010). Diagnostik der Hochbegabung im späten Jugend- und frühen Erwachsenenalter am Beispiel der Frühstudierendenprogramme. In F. Preckel, W. Schneider & H. Holling (Hrsg.), *Jahrbuch der Pädagogischen Diagnostik – Tests & Trends, Band Hochbegabung* (S. 267–292). Göttingen: Hogrefe.
- Süß, H.M. (1996). *Intelligenz, Wissen und Problemlösen: Kognitive Voraussetzungen für erfolgreiches Handeln bei computersimulierten Problemen*. Göttingen: Hogrefe.

- Süß, H.M. (2003). Intelligenztheorien. In H.J. Kubinger & A.O. Jäger (Hrsg.), *Schlüsselbegriffe der Psychologischen Diagnostik* (S. 217–224). Weinheim: Psychologie Verlags Union.
- Swanson, H.L. (2008). Working memory and intelligence in children: What develops? *Journal of Educational Psychology*, 100 (3), 581–602.
- Thomas, J., Zoelch, C., Seitz-Stein, K. & Schumann-Hengsteler, R. (2006). Phonologische und zentral-exekutive Arbeitsgedächtnisprozesse bei der mentalen Addition und Multiplikation bei Grundschulkindern. *Psychologie in Erziehung und Unterricht*, 53 (4), 275–290.
- Thorndike, E.L., Edward L. & Bregman, E.O. (1926). *The measurement of intelligence*. New York, Teachers College Bureau of Publications.
- Thorndike, R.L. & Hagen, E. (1971/1978). *Cognitive Abilities Test*. Boston: Houghton-Mifflin.
- Thurstone, L.L. (1931). Multiple factor analysis. *Psychological Review*, 38, 406–427.
- Thurstone, L.L. (1938). *Primary mental abilities*. Chicago, IL: University of Chicago Press.
- Third International Mathematics and Science Study (TIMSS) [Dritte Internationale Mathematik- und Naturwissenschaftsstudie]
- Trost, G. (1998). Identification of highly gifted adolescents: Methods and experiences. In K.A. Heller & J.F. Feldhusen (Eds.), *Identifying and nurturing the gifted. An international perspective* (S. 83–91). Toronto: Huber.
- Urbina, S. (2011). Tests of intelligence. In R.J. Sternberg & S.B. Kaufman (Eds.), *The Cambridge handbook of intelligence* (pp. 20–38). Cambridge: Cambridge University Press.
- van der Maas, H.L.J., Dolan, C.V., Grasman, R.P.P.P., Wicherts, J.M., Huizenga H.M. & Raijmakers, M.E.J. (2006). A dynamical model of general intelligence: The positive manifold of intelligence by mutualism. *Psychological Review*, 113, 842–861.
- Vernon, P.E. (1950). *The structure of human abilities*. New York: John Wiley & Sons.
- Vernon, P.E. (1969). *Intelligence and cultural environment*. London: Methuen.
- Vock, M., Preckel, F. & Holling, H. (2011). Mental abilities and school achievement: A test of a mediation hypothesis. *Intelligence*, 39, 357–369.
- Voelkle, M.C., Wittmann, W.W. & Ackerman, P.L. (2006). Abilities and skill acquisition: A latent growth curve approach. *Learning and Individual Differences*, 16, 303–319.
- Wang, M.C., Haertel, G.D. & Walberg, H.J. (1993). Toward a knowledge base for school learning. *Review of Educational Research*, 63, 249–294.
- Wasserman, J.D. (2012). A history of intelligence assessment: The unfinished tapestry. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (3rd ed., pp. 2–55). New York: Guilford Press.
- Wasserman, J.D. & Tulskey, D.S. (2005). A history of intelligence assessment. In D.P. Flanagan & P.L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests and issues* (2nd ed., pp. 3–22). New York: Guilford Press.
- Watkins, M.W., Lei, P.W. & Canivez, G.L. (2007). Psychometric intelligence and achievement: A cross-lagged panel analysis. *Intelligence*, 35, 59–68.

- Weber, H.S., Lu, L., Shi, J. & Spinath, F.M. (2013). The roles of cognitive and motivational predictors in explaining school achievement in elementary school. *Learning and Individual Differences*, 25, 85–92.
- Weber, H.M., Rücker, S., Büttner, P., Petermann, F. & Daseking, M. (2015). Zum Zusammenhang von allgemeinen kognitiven Fähigkeiten und schulischer Leistung: Welche Rolle spielt das Lernverhalten? *Gesundheitswesen*, 77 (10), 820–826.
- Weberschock, U. & Grube, D. (2006). Zur Stabilität von Einflüssen der Arbeitsgedächtniskapazität und des arithmetischen Faktenwissens auf Rechenleistungen von Viertklässlern. *Psychologie in Erziehung und Unterricht*, 53, 293–302.
- Wechsler, D. (1944). *Measurement of adult intelligence* (3rd ed.). Baltimore: Williams and Wilkins.
- Weinert, F.E. (1997). Wissen und Denken. Die unterschätzte Bedeutung des Gedächtnisses für das menschliche Denken. *Naturwissenschaftliche Rundschau*, 50 (5), 169–174.
- Weinert, F.E. (Hrsg.) (1998). *Entwicklung im Kindesalter*. Weinheim: Beltz PVU.
- Weinert, F.E. (2001a). Vergleichende Leistungsmessung in Schulen – eine umstrittene Selbstverständlichkeit. In F. E. Weinert (Hrsg.), *Leistungsmessungen in Schulen* (S. 17–31). Weinheim: Beltz.
- Weinert, F.E. (2001b). Schulleistungen – Leistungen der Schule oder der Schüler? In F.E. Weinert (Hrsg.), *Leistungsmessungen in Schulen* (S. 73–86). Weinheim: Beltz.
- Weinert, F.E. (Hrsg.) (2001). *Leistungsmessungen in Schulen*. Weinheim: Beltz.
- Weinert, F.E. & Helmke, A. (1997). *Die Münchner Grundschulstudie Scholastik. Wissenschaftliche Grundlagen, Zielsetzungen, Realisierungsbedingungen und Ergebnisperspektiven. Theoretischer Ertrag und praktischer Nutzen der Scholastik-Studie zur Entwicklung im Grundschulalter* (Forschungsbericht). München: Max-Planck-Institut für Psychologische Forschung.
- Weinert, F.E. & Helmke, A. (1993). Wie bereichsspezifisch verläuft die kognitive Entwicklung. In R. Duit & W. Gräber (Hrsg.), *Kognitive Entwicklung und Lernen in den Naturwissenschaften* (S. 27–45) Kiel: Institut für die Pädagogik der Naturwissenschaften.
- Weinert, F.E. & Helmke, A. (Hrsg.) (1997). *Entwicklung im Grundschulalter*. Weinheim: Beltz.
- Weinert, F.E. & Hany, E.A. (2000) The role of intelligence as a major determinant of a successful occupational life. In C.F.M. van Lieshout & P.G. Heymans (Eds.), *Developing talent across the life span* (pp. 67–99). Hove, UK: Psychology Press.
- Weiß, R.H. & Osterland, J. (1979). *CFT1. Grundintelligenztest Skala 1: Handanweisung für die Durchführung, Auswertung und Interpretation* (3. Aufl.). Braunschweig: Westermann.
- Weiß, R.H. (2008). *CFT 20 R. Grundintelligenztest Skala 2 – Revision*. Göttingen: Hogrefe.
- Weißhaupt S, Peucker S. & Wirtz M. (2006). Diagnose mathematischen Vorwissens im Vorschulalter und Vorhersage von Rechenleistungen und Rechenschwierigkeiten in der Grundschule. *Psychologie in Erziehung und Unterricht*, 53, 236–245.
- Widaman, K.F., Ferrer, E. & Conger, R.D. (2010). Factorial invariance within longitudinal structural equation models. Measuring the same construct across time. *Child Development Perspectives*, 4 (1), 10–18.

- Wiesner, M. & Schanding, G.T. (2013). Exploratory structural equation modeling, bifactor models, and standard confirmatory factor analysis models: Application to the BASC-2 Behavioral and Emotional Screening System Teacher Form. *Journal of School Psychology*, 51 (6), 751–763.
- Willis, J.O., Dumont, R., & Kaufman, A.S. (2011). Factor–analytic models of intelligence. In R. J. Sternberg & S. B. Kaufman (Eds.), *The Cambridge handbook of intelligence* (pp. 39–57). New York: Cambridge University Press.
- Winship, C. & Korenman, S. (1997). Does staying in school make you smarter? The effect of education on IQ in *The Bell Curve*. In B. Devlin, S.E. Fienberg, D.P. Resnick & K. Roeder (Eds.), *Intelligence, genes, and success: Scientists respond to the Bell Curve* (pp. 215–234). New York: Springer-Verlag.
- Wittman, M., Eisenkolb, A. & Perleth, Ch. (1997). *Neue Intelligenztests. Ein umfassendes Test- und Übungsprogramm*. Augsburg: Augustus.
- Yousfi, S. & Steyer, R. (2006). Klassische Testtheorie. In F. Petermann & M. Eid (Hrsg.), *Handbuch der Psychologischen Diagnostik* (S. 288–303). Göttingen: Hogrefe.
- Yen, C.J., Konold, T.R. & McDermott, P.A. (2004). Does learning behavior augment cognitive ability as an indicator of academic achievement? *Journal of School Psychology*, 42 (2), 157–169.
- Ziegler, A. (2004). Stabilität von Intelligenz und Hochbegabung im Vorschulalter. *LVH aktuell*, 13, 3–8.
- Ziegler, A. (2005). The Actiotope Model of Giftedness. In R. J. Sternberg & J. E. Davidson (Eds.), *Conceptions of giftedness* (2nd ed., pp. 411–436). New York: Cambridge University Press.

6.2 Abbildungsverzeichnis

Abbildung 1:	Spearman's g	25
Abbildung 2:	Cattell's investment theory	28
Abbildung 3:	CHC abilities as parameters of information processing	33
Abbildung 4:	Münchner Hochbegabungsmodell	43
Abbildung 5:	Das Münchner dynamische Begabungs-Leistungs-Modell (MDAAM)	44
Abbildung 6:	Beispielaufgabe DEMAT	81
Abbildung 7:	Beispielaufgabe KFT	82
Abbildung 8:	Beispielaufgabe DEMAT 1	82
Abbildung 9:	Beispielaufgabe DEMAT 1	83
Abbildung 10:	Übersicht über die durchgeführten Datenerhebungen von 2006 bis 2009	163
Abbildung 11:	Beispiel Matrizen N2	166
Abbildung 12:	Latent-State-Model – Messmodell	176
Abbildung 13:	Latent-State-Model I – Strukturmodell	176
Abbildung 14:	Latent-State-Model II – Strukturmodell	177
Abbildung 15:	Autoregressives Modell 1. Ordnung	177
Abbildung 16:	Autoregressives Modell 2. Ordnung	177
Abbildung 17:	Autoregressives Modell 3. Ordnung	181
Abbildung 18:	Crossed-Lagged-Modell 4. Ordnung	182
Abbildung 19:	Crossed-Lagged-Modell 5. Ordnung	183
Abbildung 20:	Crossed-Lagged-Modell 6. Ordnung	184
Abbildung 21:	Basis-Messmodelle beider Konstrukte	185
Abbildung 22:	Faktorielle Differenziertheit KFT 1 vs. DEMAT 1	186
Abbildung 23:	Faktorielle Differenziertheit KFT 2 vs. DEMAT 2	187
Abbildung 24:	Faktorielle Differenziertheit KFT 3 vs. DEMAT 3	188
Abbildung 25:	Faktorielle Differenziertheit KFT 4 vs. DEMAT 4	189
Abbildung 26:	Verteilung der Rohwerte zum <i>ersten</i> Messzeitpunkt im KFT	193
Abbildung 27:	Verteilung der Rohwerte zum <i>zweiten</i> Messzeitpunkt im KFT	194
Abbildung 28:	Verteilung der Rohwerte zum <i>dritten</i> Messzeitpunkt im KFT	195
Abbildung 29:	Verteilung der Rohwerte zum <i>vierten</i> Messzeitpunkt im KFT	196
Abbildung 30:	Verteilung der Rohwerte im DEMAT (Gesamtwerte)	198

Abbildung 31:	Strukturmodell zur Testung der Invarianz.....	208
Abbildung 32:	Prüfung der Messinvarianz – Stufe I; DEMAT-Residualkorrelationen	209
Abbildung 33:	Prüfung der Messinvarianz – Stufe II; mit signifikanten DEMAT-Residualkorrelationen	210
Abbildung 34:	Autoregressives Modell 1. Ordnung	211
Abbildung 35:	Autoregressives Modell 2. Ordnung	225
Abbildung 36:	Autoregressives Modell 3. Ordnung	235
Abbildung 37:	Crossed-Lagged-Effekte 1. Ordnung (Modell 4)	236
Abbildung 38:	Crossed-Lagged-Effekte 2. Ordnung (Modell 5)	244
Abbildung 39:	Crossed-Lagged-Effekte 3. Ordnung (Modell 6)	245
Abbildung 40:	KFT-Subtestkorrelationen; Residualkorrelationen auf latenter Ebene	246
Abbildung 41:	DEMAT-Subtestkorrelationen; Residualkorrelationen auf latenter Ebene	250
Abbildung 42:	komplette Darstellung der berücksichtigten Parameter im Crossed-Lagged-Modell 6	252
Abbildung 43:	Faktorielle Differenziertheit KFT 1 vs. DEMAT 1.....	255
Abbildung 44:	Faktorielle Differenziertheit KFT 2 vs. DEMAT 2.....	258
Abbildung 45:	Faktorielle Differenziertheit KFT 3 vs. DEMAT 3.....	258
Abbildung 46:	Faktorielle Differenziertheit KFT 4 vs. DEMAT 4.....	259

6.3 Tabellenverzeichnis

Tabelle 1:	Beschreibung der Gesamtstichprobe nach Geschlecht für die Regionen Rostock <i>und</i> Rügen.....	158
Tabelle 2:	Beschreibung der Gesamtstichprobe nach Geschlecht und Region für die Region Rostock	159
Tabelle 3:	Beschreibung der Gesamtstichprobe nach Geschlecht und Region für die Region Rügen.....	160
Tabelle 4:	Übersicht über die den einzelnen Fragestellungen zugrundeliegenden Fallzahlen	162
Tabelle 5:	Übersicht über Skalen und Items zur Messung mathematischer Schulleistungen aus der Reihe Deutscher Mathematiktests (DEMAT) 1. bis 4. Klassenstufe	171
Tabelle 6:	Übersicht über die Teilstichproben zur Beantwortung der Fragestellungen	174
Tabelle 7:	Kognitive Fähigkeiten (KFT) für die vier Messzeitpunkte.....	191
Tabelle 8:	Tests auf Normalverteilung nach Kolmogorov-Smirnov mit Signifikanzprüfung	192
Tabelle 9:	Deutscher Mathematiktest (DEMAT) für die vier Messzeitpunkte.....	197
Tabelle 10:	Korrelationen über die vier Messzeitpunkte – manifeste Variablen I	199
Tabelle 11:	Korrelationen über die vier Messzeitpunkte – manifeste Variablen II	200
Tabelle 12:	Korrelationen zwischen den manifesten Variablen (standardisiert), DEMAT – Subskala „Geometrie“	201
Tabelle 13:	Korrelationen zwischen den manifesten Variablen (standardisiert), DEMAT – Subskala „Arithmetik“	202
Tabelle 14:	Korrelationen zwischen den manifesten Variablen (standardisiert), DEMAT – Subskala „Sachrechnen“	203
Tabelle 15:	Korrelationen zwischen den manifesten Variablen (standardisiert), KFT – Subskala „quantitativ“	203
Tabelle 16:	Korrelationen zwischen den manifesten Variablen (standardisiert), KFT – Subskala „verbal“	204
Tabelle 17:	Korrelationen zwischen den manifesten Variablen (standardisiert), KFT – Subskala „nonverbal“	205
Tabelle 18:	Effektstärken und Faktorreliabilitäten (KFT) für vier Messzeitpunkte ...	206

Tabelle 19:	Modellvergleich Invarianz KFT, Passung der einzelnen Modelle zueinander	207
Tabelle 20:	Modellvergleich χ^2 -Differenzentest KFT, Beurteilung der Passung der Modelle zueinander.....	213
Tabelle 21:	Passungswerte zur Prüfung der Invarianz (jeweils ohne und mit Kontrolle der Mehrebenenstruktur)	226
Tabelle 22:	Effektstärken und Faktorreliabilitäten (DEMAT) für vier Messzeitpunkte	231
Tabelle 23:	χ^2 -Differenzentest DEMAT-Modellvergleich.....	232
Tabelle 24:	Passung der einzelnen Modelle für DEMAT	233
Tabelle 25:	Prüfung der Invarianz (jeweils ohne und mit Kontrolle der Mehr- ebenenstruktur)	234
Tabelle 26:	Passungswerte für autoregressive Modelle – Stabilität	240
Tabelle 27:	Passungswerte angewandter Modelle zur Prüfung autoregressiver Effekte (Modelle 1 bis 3) jeweils ohne und mit Kontrolle der Mehr- ebenenstruktur	241
Tabelle 28:	Passungswerte angewandter Modelle – autoregressiv (1 bis 3) und Crossed Lagged (4 bis 6)	242
Tabelle 29:	Regressionen aller sechs Modelle – autoregressiv (1 bis 3) und Crossed Lagged (4 bis 6)	248
Tabelle 30:	Passungswerte angewandter Modelle – Passungswerte angewandter Modelle zur Prüfung der kreuzverzögerten Effekte (Modelle 4-6) mit und ohne Mehrebenenkontrolle	249
Tabelle 31:	Korrelationen zum ersten und zweiten Messzeitpunkt – manifeste Faktoren	261
Tabelle 32:	Korrelationen zum dritten und vierten Messzeitpunkt – manifeste Faktoren	262
Tabelle 33:	Passungswerte angewandter Modelle zur Prüfung der Unabhängigkeit der Konstrukte der 4 Messzeitpunkte jeweils ohne und mit Kontrolle der Mehrebenenstruktur	263
Tabelle 34:	Standardisierte Residualkorrelationen für KFT „verbal“	302
Tabelle 35:	Standardisierte Residualkorrelationen für KFT „quantitativ“	302
Tabelle 36:	Standardisierte Residualkorrelationen für KFT „nonverbal“	303

Tabelle 37:	Faktorladungen und Faktorreliabilität KFT 1-4 mit Subskalen „verbal“, „quantitativ“ und „nonverbal“ für die vier Messzeitpunkte konfiguraler Invarianz	303
Tabelle 38:	Residualvarianzen und quadrierte Faktorladungen KFT 1-4 mit Subskalen „verbal“, „quantitativ“ und „nonverbal“ für die vier Messzeitpunkte mit Residualkovarianzen	304
Tabelle 39:	Korrelationen zwischen den latenten Variablen (standardisiert), Latent-State-Modell 1a – konfigurale Invarianz mit Residualkovarianzen	305
Tabelle 40:	Standardisierte Residualkorrelationen für KFT „verbal“, Latent-State-Modell 2a – metrische Invarianz mit Residualkovarianzen	306
Tabelle 41:	Standardisierte Residualkorrelationen für KFT „quantitativ“, Latent-State-Modell 2a – metrische Invarianz mit Residualkovarianzen	306
Tabelle 42:	Standardisierte Residualkorrelationen für KFT „nonverbal“, Latent-State-Modell 2a – metrische Invarianz mit Residualkovarianzen	306
Tabelle 43:	Faktorladungen und Effektstärken KFT 1-4 mit Subskalen „verbal“, „quantitativ“ und „nonverbal“ für die vier Messzeitpunkte, Latent-State-Modell 2a – metrische Invarianz mit Residualkovarianzen	307
Tabelle 44:	Korrelationen zwischen den latenten Variablen (standardisiert), Latent-State-Modell 2a – metrische Invarianz mit Residualkovarianzen	307
Tabelle 45:	Residualvarianzen und quadrierte Faktorladungen KFT 1-4 mit Subskalen „verbal“, „quantitativ“ und „nonverbal“ für die vier Messzeitpunkte, Latent-State-Modell 2a – metrische Invarianz mit Residualkovarianzen	308
Tabelle 46:	Standardisierte Residualkorrelationen für KFT „verbal“, Latent-State-Modell 3a – skalare Invarianz mit Residualkovarianzen	309
Tabelle 47:	Standardisierte Residualkorrelationen für KFT „quantitativ“, Latent-State-Modell 3a – skalare Invarianz mit Residualkovarianzen	309

Tabelle 48:	Standardisierte Residualkorrelationen für KFT „nonverbal“, Latent-State-Modell 3a – skalare Invarianz mit Residualkovarianzen	309
Tabelle 49:	Korrelationen zwischen den latenten Variablen (standardisiert), Latent-State-Modell 3a – skalare Invarianz mit Residualkovarianzen	310
Tabelle 50:	Residualvarianzen und quadrierte Faktorladungen KFT 1-4 mit Subskalen „verbal“, „quantitativ“ und „nonverbal“ für die vier Messzeitpunkte, Latent-State-Modell 3a – skalare Invarianz mit Residualkovarianzen	310
Tabelle 51:	Faktorladungen DEMAT 1-4 mit Subskalen „Arithmetik“, „Sachrechnen“ und „Geometrie“ für die vier Messzeitpunkte	311
Tabelle 52:	Standardisierte Residualkorrelationen für DEMAT „Arithmetik“, Latent-State-Modell 1a – konfigurale Invarianz	312
Tabelle 53:	Standardisierte Residualkorrelationen für DEMAT „Geometrie“, Latent-State-Modell 1a – konfigurale Invarianz	312
Tabelle 54:	Standardisierte Residualkorrelationen für DEMAT „Sachrechnen“, Latent-State-Modell 1a – konfigurale Invarianz	312
Tabelle 55:	Residualvarianzen und quadrierte Faktorladungen DEMAT 1-4 mit Subskalen „Arithmetik“, „Sachrechnen“ und „Geometrie“ für die vier Messzeitpunkte, Latent-State-Modell 1b – konfigurale Invarianz mit 4 Residualkovarianzen	313
Tabelle 56:	Faktorladungen und Faktorreliabilität der Subskalen, Latent-State-Modell 1b – konfigurale Invarianz mit 4 Residualkovarianzen	314
Tabelle 57:	Faktorladungen der Subskalen (standardisiert), Latent-State-Modell 2b – metrische Invarianz mit 4 Residualkovarianzen	315
Tabelle 58:	Korrelationen zwischen den latenten Variablen (standardisiert), Latent-State Modell 2b – metrische Invarianz mit 4 Residualkovarianzen	316
Tabelle 59:	Korrelationen zwischen den latenten Variablen (standardisiert), Latent-State-Modell 2c – partiell-metrische Invarianz mit 4 Residualkovarianzen	317

Tabelle 60:	Faktorladungen der Subskalen (standardisiert), Latent-State-Modell 2c – partiell-metrische Invarianz mit 4 Residualkovarianzen	317
Tabelle 61:	Faktorladungen der Subskalen (standardisiert), Latent-State-Modell 3b – skalare Invarianz mit 4 Residualkovarianzen	318
Tabelle 62:	Passungswerte für getestete Modelle, autoregressiv (3) und Crossed Lagged (4, 6)	319

7 Anhang

7.1 KFT Invarianzprüfung

Modell 1a mit konfiguraler Invarianz

Tabelle 34

Standardisierte Residualkorrelationen für KFT „verbal“, Latent-State-Modell 1a – konfigurale Invarianz mit Residualkovarianzen

	KFT 2	KFT 3	KFT 4
KFT 1	.47	.20	.18
KFT 2		.23	.26
KFT 3			.52

Tabelle 35

Standardisierte Residualkorrelationen für KFT „quantitativ“, Latent-State-Modell 1a – konfigurale Invarianz mit Residualkovarianzen

	KFT 2	KFT 3	KFT 4
KFT 1	- .45	- .18	- .20
KFT 2		- .26	-.17
KFT 3			.12

Tabelle 36

Standardisierte Residualkorrelationen für KFT „nonverbal“, Latent-State-Modell 1a konfigurale Invarianz mit Residualkovarianzen

	KFT 2	KFT 3	KFT 4
KFT 1	.31	.11	.15
KFT 2		0,32	.21
KFT 3			.33

Tabelle 37

Faktorladungen und Faktorreliabilität KFT 1-4 mit Subskalen „verbal“, „quantitativ“ und „nonverbal“ für die vier Messzeitpunkte konfiguraler Invarianz

	Faktorladung λ	Faktorreliabilität ω
Messzeitpunkt 1		
KFT 1 by KFT 1 verbal	.69	
KFT 1 by KFT 1 quantitativ	.92	.80
KFT 1 by KFT 1 nonverbal	.63	
Messzeitpunkt 2		
KFT 2 by KFT 2 verbal	.73	
KFT 2 by KFT 2 quantitativ	.95	.84
KFT 2 by KFT 2 nonverbal	.71	
Messzeitpunkt 3		
KFT 3 by KFT 3 verbal	.77	
KFT 3 by KFT 3 quantitativ	.86	.83
KFT 3 by KFT 3 nonverbal	.74	
Messzeitpunkt 4		
KFT 4 by KFT 4 verbal	.81	
KFT 4 by KFT 4 quantitativ	.84	.85
KFT 4 by KFT 4 nonverbal	.76	

Tabelle 38

Residualvarianzen und quadrierte Faktorladungen KFT 1-4 mit Subskalen „verbal“, „quantitativ“ und „nonverbal“ für die vier Messzeitpunkte mit Residualkovarianzen

	<i>Residualvarianz ε (standardisiert)</i>	<i>quadrierte Faktorladung λ^2</i>
Messzeitpunkt 1		
KFT 1 verbal	.52	.48
KFT 1 quantitativ	.15	.85
KFT 1 nonverbal	.60	.40
Messzeitpunkt 2		
KFT 2 verbal	.47	.53
KFT 2 quantitativ	.10	.90
KFT 2 nonverbal	.49	.51
Messzeitpunkt 3		
KFT 3 verbal	.41	.59
KFT 3 quantitativ	.26	.74
KFT 3 nonverbal	.46	.54
Messzeitpunkt 4		
KFT 4 verbal	.34	.66
KFT 4 quantitativ	.29	.71
KFT 4 nonverbal	0,43	.58

Tabelle 39

Korrelationen zwischen den latenten Variablen (standardisiert), Latent-State-Modell 1a – konfigurale Invarianz mit Residualkovarianzen

	KFT 1	KFT 2	KFT 3	KFT 4
KFT 1	1.00			
KFT 2	.82	1.00		
KFT 3	.74	.81	1.00	
KFT 4	.73	.76	.87	1.00

Modell 2a mit metrischer Invarianz

Tabelle 40

Standardisierte Residualkorrelationen für KFT „verbal“, Latent-State-Modell 2a – metrische Invarianz mit Residualkovarianzen

	KFT 2	KFT 3	KFT 4
KFT 1	.48	.18	.11
KFT 2		.19	.16
KFT 3			.54

Tabelle 41

Standardisierte Residualkorrelationen für KFT „quantitativ“, Latent-State-Modell 2a – metrische Invarianz mit Residualkovarianzen

	KFT 2	KFT 3	KFT 4
KFT 1	-.06	-.14	-.28
KFT 2		-.21	-.30
KFT 3			.01

Tabelle 42

Standardisierte Residualkorrelationen für KFT „nonverbal“, Latent-State-Modell 2a – metrische Invarianz mit Residualkovarianzen

	KFT 2	KFT 3	KFT 4
KFT 1	.27	.09	.15
KFT 2		.31	.24
KFT 3			.36

Tabelle 43

Faktorladungen und Effektstärken KFT 1-4 mit Subskalen „verbal“, „quantitativ“ und „nonverbal“ für die vier Messzeitpunkte, Latent-State-Modell 2a – metrische Invarianz mit Residualkovarianzen

	<i>Faktorladung λ</i>	<i>Effektstärke ω</i>
Messzeitpunkt 1		
KFT 1 by KFT 1 verbal	.73	
KFT 1 by KFT 1 quantitativ	.86	.80
KFT 1 by KFT 1 nonverbal	.69	
Messzeitpunkt 2		
KFT 2 by KFT 2 verbal	.78	
KFT 2 by KFT 2 quantitativ	.91	.84
KFT 2 by KFT 2 nonverbal	.71	
Messzeitpunkt 3		
KFT 3 by KFT 3 verbal	.71	
KFT 3 by KFT 3 quantitativ	.88	.83
KFT 3 by KFT 3 nonverbal	.72	
Messzeitpunkt 4		
KFT 4 by KFT 4 verbal	.63	
KFT 4 by KFT 4 quantitativ	.91	.85
KFT 4 by KFT 4 nonverbal	.73	

Tabelle 44

Korrelationen zwischen den latenten Variablen (standardisiert), Latent-State Modell 2a – metrische Invarianz mit Residualkovarianzen

	KFT 1	KFT 2	KFT 3
KFT 2	.84	1.00	
KFT 3	.75	.81	1.00
KFT 4	.73	.76	.85

Tabelle 45

Residualvarianzen und quadrierte Faktorladungen KFT 1-4 mit Subskalen „verbal“, „quantitativ“ und „nonverbal“ für die vier Messzeitpunkte, Latent-State-Modell 2a – metrische Invarianz mit Residualkovarianzen

	<i>Residualvarianz ε (standardisiert)</i>	<i>λ^2</i>
Messzeitpunkt 1		
KFT 1 verbal	.47	.53
KFT 1 quantitativ	.26	.74
KFT 1 nonverbal	.52	.47
Messzeitpunkt 2		
KFT 2 verbal	.39	.61
KFT 2 quantitativ	.17	.83
KFT 2 nonverbal	.49	.51
Messzeitpunkt 3		
KFT 3 verbal	.50	.50
KFT 3 quantitativ	.23	.78
KFT 3 nonverbal	.48	.52
Messzeitpunkt 4		
KFT 4 verbal	.61	.40
KFT 4 quantitativ	.17	.83
KFT 4 nonverbal	.46	.54

KFT Modell 3a mit skalarer Invarianz*Tabelle 46*

Standardisierte Residualkorrelationen für KFT „verbal“, Latent-State-Modell 3a – skalare Invarianz mit Residualkovarianzen

	KFT 2	KFT 3	KFT 4
KFT 1	.45	- .24	.29
KFT 2		.09	.20
KFT 3			- .13

Tabelle 47

Standardisierte Residualkorrelationen für KFT „quantitativ“, Latent-State-Modell 3a – skalare Invarianz mit Residualkovarianzen

	KFT 2	KFT 3	KFT 4
KFT 1	- .02	- .01	- .23
KFT 2		- .11	- .40
KFT 3			- .08

Tabelle 48

Standardisierte Residualkorrelationen für KFT „nonverbal“, Latent-State-Modell 3a – skalare Invarianz mit Residualkovarianzen

	KFT 2	KFT 3	KFT 4
KFT 1	.50	- .35	- .20
KFT 2		- .38	- .26
KFT 3			.66

Tabelle 49

Korrelationen zwischen den latenten Variablen (standardisiert), Latent-State-Modell 3a – skalare Invarianz mit Residualkovarianzen

	KFT 1	KFT 2	KFT 3
KFT 2	.84	1.00	
KFT 3	.73	.78	1.00
KFT 4	.74	.78	.86

Tabelle 50

Residualvarianzen und quadrierte Faktorladungen KFT 1-4 mit Subskalen „verbal“, „quantitativ“ und „nonverbal“ für die vier Messzeitpunkte, Latent-State-Modell 3a – skalare Invarianz mit Residualkovarianzen

	Residualvarianz ε (standardisiert)	λ^2
Messzeitpunkt 1		
KFT 1 verbal	.50	.50
KFT 1 quantitativ	.28	.72
KFT 1 nonverbal	.58	.42
Messzeitpunkt 2		
KFT 2 verbal	.38	.61
KFT 2 quantitativ	.21	.79
KFT 2 nonverbal	.62	.38
Messzeitpunkt 3		
KFT 3 verbal	.75	.25
KFT 3 quantitativ	.20	.80
KFT 3 nonverbal	.70	.30
Messzeitpunkt 4		
KFT 4 verbal	.65	.35
KFT 4 quantitativ	.21	.79
KFT 4 nonverbal	.57	.43

7.2 Invarianzprüfung DEMAT

DEMAT Modell 1a konfiguraler Invarianz

Tabelle 51

Faktorladungen DEMAT 1-4 mit Subskalen „Arithmetik“, „Sachrechnen“ und „Geometrie“ für die vier Messzeitpunkte

	<i>Faktorladung λ</i>
Messzeitpunkt 1	
DEMAT 1 by DEMAT 1 Arithmetik	.85
DEMAT 1 by DEMAT 1 Sachrechnen	.73
DEMAT 1 by DEMAT 1 Geometrie	.42
Messzeitpunkt 2	
DEMAT 2 by DEMAT 2 Arithmetik	.91
DEMAT 2 by DEMAT 2 Sachrechnen	.73
DEMAT 2 by DEMAT 2 Geometrie	.43
Messzeitpunkt 3	
DEMAT 3 by DEMAT 3 Arithmetik	.76
DEMAT 3 by DEMAT 3 Sachrechnen	.64
DEMAT 3 by DEMAT 3 Geometrie	.62
Messzeitpunkt 4	
DEMAT 4 by DEMAT 4 Arithmetik	.70
DEMAT 4 by DEMAT 4 Sachrechnen	.77
DEMAT 4 by DEMAT 4 Geometrie	.59

DEMAT Modell 1a mit konfiguraler Invarianz

Tabelle 52

Standardisierte Residualkorrelationen für DEMAT „Arithmetik“, Latent-State-Modell 1a – konfigurale Invarianz

	DEMAT 2	DEMAT 3	DEMAT 4
DEMAT 1	.11	.07	.11
DEMAT 2		.14	.00
DEMAT 3			.20

Tabelle 53

Standardisierte Residualkorrelationen für DEMAT „Geometrie“, Latent-State-Modell 1a – konfigurale Invarianz

	DEMAT 2	DEMAT 3	DEMAT 4
DEMAT 1	.05	.02	- .02
DEMAT 2		.15	.14
DEMAT 3			.26

Tabelle 54

Standardisierte Residualkorrelationen für DEMAT „Sachrechnen“, Latent-State-Modell 1a – konfigurale Invarianz

	DEMAT 2	DEMAT 3	DEMAT 4
DEMAT 1	.01	- .02	- .03
DEMAT 2		.01	.01
DEMAT 3			- .04

Tabelle 55

Residualvarianzen und quadrierte Faktorladungen DEMAT 1-4 mit Subskalen „Arithmetik“, „Sachrechnen“ und „Geometrie“ für die vier Messzeitpunkte, Latent-State-Modell 1b – konfigurale Invarianz mit 4 Residualkovarianzen

	<i>Residualvarianz ε (standardisiert)</i>	<i>λ^2</i>
Messzeitpunkt 1		
DEMAT 1 Arithmetik	.26	.74
DEMAT 1 Sachrechnen	.49	.52
DEMAT 1 Geometrie	.83	.,17
Messzeitpunkt 2		
DEMAT 2 Arithmetik	.16	.84
DEMAT 2 Sachrechnen	.48	.53
DEMAT 2 Geometrie	.82	.18
Messzeitpunkt 3		
DEMAT 3 Arithmetik	.39	.61
DEMAT 3 Sachrechnen	.62	.39
DEMAT 3 Geometrie	.63	.37
Messzeitpunkt 4		
DEMAT 4 Arithmetik	.51	.49
DEMAT 4 Sachrechnen	.42	.58
DEMAT 4 Geometrie	.65	.35

Tabelle 56

Faktorladungen und Faktorreliabilität der Subskalen, Latent-State-Modell 1b – konfigurale Invarianz mit 4 Residualkovarianzen

	<i>Faktorladung λ</i>	<i>Faktorreliabilität ω</i>
Messzeitpunkt 1		
DEMAT 1 by DEMAT 1 Arithmetik	.86	
DEMAT 1 by DEMAT 1 Sachrechnen	.72	.72
DEMAT 1 by DEMAT 1 Geometrie	.42	
Messzeitpunkt 2		
DEMAT 2 by DEMAT 2 Arithmetik	.92	
DEMAT 2 by DEMAT 2 Sachrechnen	.73	.75
DEMAT 2 by DEMAT 2 Geometrie	.43	
Messzeitpunkt 3		
DEMAT 3 by DEMAT 3 Arithmetik	.78	
DEMAT 3 by DEMAT 3 Sachrechnen	.62	.71
DEMAT 3 by DEMAT 3 Geometrie	.61	
Messzeitpunkt 4		
DEMAT 4 by DEMAT 4 Arithmetik	.70	
DEMAT 4 by DEMAT 4 Sachrechnen	.76	.73
DEMAT 4 by DEMAT 4 Geometrie	.59	

DEMAT Modell 2b mit metrischer Invarianz*Tabelle 57*

Faktorladungen der Subskalen (standardisiert), Latent-State-Modell 2b – metrische Invarianz mit 4 Residualkovarianzen

	<i>Faktorladung λ</i>
Messzeitpunkt 1	
DEMAT 1 by DEMAT 1 Arithmetik	.82
DEMAT 1 by DEMAT 1 Sachrechnen	.72
DEMAT 1 by DEMAT 1 Geometrie	.53
Messzeitpunkt 2	
DEMAT 2 by DEMAT 2 Arithmetik	.88
DEMAT 2 by DEMAT 2 Sachrechnen	.78
DEMAT 2 by DEMAT 2 Geometrie	.33
Messzeitpunkt 3	
DEMAT 3 by DEMAT 3 Arithmetik	.82
DEMAT 3 by DEMAT 3 Sachrechnen	.58
DEMAT 3 by DEMAT 3 Geometrie	.17
Messzeitpunkt 4	
DEMAT 4 by DEMAT 4 Arithmetik	.82
DEMAT 4 by DEMAT 4 Sachrechnen	.52
DEMAT 4 by DEMAT 4 Geometrie	.19

Tabelle 58

Korrelationen zwischen den latenten Variablen (standardisiert), Latent-State-Modell 2b – metrische Invarianz mit 4 Residualkovarianzen

	DEMAT 1	DEMAT 2	DEMAT 3
DEMAT 2	.80	1.00	
DEMAT 3	.81	.85	1.00
DEMAT 4	.79	.84	.94

DEMAT Modell 2c mit partiell metrischer Invarianz*Tabelle 59*

Korrelationen zwischen den latenten Variablen (standardisiert), Latent-State-Modell 2c – partiell-metrische Invarianz mit 4 Residualkovarianzen

	DEMAT 1	DEMAT 2	DEMAT 3
DEMAT 2	.78	1.00	
DEMAT 3	.78	.84	1.00
DEMAT 4	.79	.87	.93

Tabelle 60

Faktorladungen der Subskalen (standardisiert), Latent-State-Modell 2c – partiell-metrische Invarianz mit 4 Residualkovarianzen

	<i>Faktorladung λ</i>
Messzeitpunkt 1	
DEMAT 1 by DEMAT 1 Arithmetik	.86
DEMAT 1 by DEMAT 1 Sachrechnen	.72
DEMAT 1 by DEMAT 1 Geometrie	.42
Messzeitpunkt 2	
DEMAT 2 by DEMAT 2 Arithmetik	.92
DEMAT 2 by DEMAT 2 Sachrechnen	.73
DEMAT 2 by DEMAT 2 Geometrie	.43
Messzeitpunkt 3	
DEMAT 3 by DEMAT 3 Arithmetik	.78
DEMAT 3 by DEMAT 3 Sachrechnen	.62
DEMAT 3 by DEMAT 3 Geometrie	.62
Messzeitpunkt 4	
DEMAT 4 by DEMAT 4 Arithmetik	.70
DEMAT 4 by DEMAT 4 Sachrechnen	.76
DEMAT 4 by DEMAT 4 Geometrie	.59

DEMAT Modell 3b mit skalarer Invarianz*Tabelle 61*

Faktorladungen der Subskalen (standardisiert), Latent-State-Modell 3b – skalare Invarianz mit 4 Residualkovarianzen

	<i>Faktorladung λ</i>
Messzeitpunkt 1	
DEMAT 1 by DEMAT 1 Arithmetik	.86
DEMAT 1 by DEMAT 1 Sachrechnen	.72
DEMAT 1 by DEMAT 1 Geometrie	.42
Messzeitpunkt 2	
DEMAT 2 by DEMAT 2 Arithmetik	.92
DEMAT 2 by DEMAT 2 Sachrechnen	.73
DEMAT 2 by DEMAT 2 Geometrie	.43
Messzeitpunkt 3	
DEMAT 3 by DEMAT 3 Arithmetik	.78
DEMAT 3 by DEMAT 3 Sachrechnen	.65
DEMAT 3 by DEMAT 3 Geometrie	.60
Messzeitpunkt 4	
DEMAT 4 by DEMAT 4 Arithmetik	.73
DEMAT 4 by DEMAT 4 Sachrechnen	.71
DEMAT 4 by DEMAT 4 Geometrie	.61

Passungswerte der drei wesentlichen vorgestellten Modelle

Tabelle 62

Passungswerte für getestete Modelle, autoregressiv (3) und Crossed Lagged (4, 6)

	ARM 3	CL 4	CL 6
χ^2	1043.82	953.57	593.90
df	216	213	208
<i>CFI</i>	.95	.95	.98
<i>RMSEA</i>	.05	.05	.03
$\Delta\chi^2$		90.25	359.67
Δdf		3	5

χ^2 = Diskrepanz; $\Delta\chi^2$ = Differenz Diskrepanz; df = Freiheitsgrade; Δdf = Differenz Freiheitsgrade; *CFI* = Comparative Fit Index; *RMSEA* = Root Mean Square Error of Approximation; *ARM* = autoregressives Modell; *CL* = Crossed Lagged

7.3 Publikationsliste

- Hoese, D. (2010). *Ein Elternratgeber mit Enrichmentvorschlägen zur Förderung begabter Kinder*. Universität Rostock.
- Hoese, D. & Perleth, Ch. (2011). *Evaluation der Weiterbildung zum Begabtenpädagogen im Elementar- und Grundschulbereich*. Evaluationsbericht qualitativer Interviews einer zweijährigen Weiterbildung. Karg Stiftung: Frankfurt.
- Perleth, C. & Hoese, D. (2012). *Die Begabungspsychologische Beratungsstelle „Odysseus-Projekt“ der Universität Rostock feierte ihr 10jähriges Bestehen*. Arbeitskreis Bildung Begabung - ABB-Information. Jahresheft 2012, 2012.
- Perleth, C., Joswig, H. & Hoese, D. (2012). *Beratungspraxis der Begabungspsychologischen Beratungsstelle des Odysseus-Projektes am Institut für Pädagogische Psychologie „Rosa und David Katz“ der Universität Rostock*. In A. Ziegler, R. Grassinger & B. Harder (Hrsg.), *Konzepte der Hochbegabtenberatung in der Praxis*. Berlin: LIT.
- Hoese, D. (2013, 08.09.2013). „*Schach und Schule*“. Vortrag zum Forschungsstand und Konzeptvorstellung des Projekts in Grundschule „Deutsche Schachschule“. Schwerin-Lankow.
- Hoese, D. (2013, 10.10.2013). „*Schach und Schule*“. Vortrag zum Forschungsstand und Konzeptvorstellung des Projekts in Grundschule John Brinkmann. Schwerin.
- Hoese, D. (2014, 08.07.2014). „*Schach und Schule*“. Vortrag zum Forschungsstand und Konzeptvorstellung des Projekts in Hamburg – Schachschule Hamburg.
- Hoese, D. (2015, 12.07.2015). „*Evaluationsbericht - Studienprojekt - Training von Selbstkompetenzen im Jugendalter*“. Vortrag im Rahmen der Studien- und Karriereorientierung für hochbegabte Jugendliche. CJD Christophorus Gymnasium Rostock.
- Hoese, D., Hildebrandt, A. & Perleth, Ch. (2015, 10.09.2015). „*Kognitive Fähigkeiten und Mathematikleistungen im Grundschulalter – Kreuzverzögerte Effekte über vier Messzeitpunkte*“. Tagungsbeitrag 5. Münsterscher Bildungskongress „Potenzialentwicklung - Begabungsförderung - Bildung der Vielfalt“ des Internationalen Centrums für Begabungsforschung (icbf) 09.-12.09.2015, Westfälische Wilhelms-Universität Münster.
- Hoese, D., Hildebrandt, A. & Perleth, Ch. (2015, 15.09.2015). „*Kognitive Fähigkeiten und Mathematikleistungen im Grundschulalter – Kreuzverzögerte Effekte über vier Messzeitpunkte*“. Poster. Hoese, D., Hildebrandt, A. & Perleth, Ch. (2015). Tagungsbeitrag 15. Fachgruppe Pädagogische Psychologie der Deutschen Gesellschaft für Psychologie (PAEPS) 14. -16. 09. 2015, Kassel.
- Hoese, D, Hildebrandt, A. & Perleth, Ch. & Pomowski, P. (2016, 08.04.2016). *Kognitive Fähigkeiten und Schach im Grundschulalter*. Posterbeitrag, Universität Rostock.
- Hoese, D, Hildebrandt, A. & Perleth, Ch. (2016, 03.03.2016). *Analysis of the Stability of Cognitive Achievements in Primary School Children using Structural Equation Modeling*. Symposium 15. International Conference European Council for High Ability (ECHA) – Talents in Motion, 02.-05. 03.2016, Wien.

- Hoese, D. (2016, 11.07.2016). „Schach und Schule“ Qualitative Ergebnisse explorativer Interviews. *Forschungsvortrag des Evaluationsprojekts, Deutsche Schachschule Schwerin*.
- Hoese, D. (2016, 12.07.2015). *Erste Forschungsergebnisse des Studienprojekts „Training von Selbstkompetenzen im Jugendalter“*. CJD Christophorus Gymnasium Rostock, 2016.
- Hoese, D, Hildebrandt, A. & Perleth, Ch. (2016, 19.09.2016). *Sind kreuzverzögerte Effekte kognitiver Fähigkeiten und der Mathematikleistung im Grundschulalter stabil über Schulen hinweg?* 50. Kongress der Deutschen Gesellschaft für Psychologie, 18.-22. Sept. 2016, Leipzig.
- Hoese, D. (2016, 05.10.2016). (Vor-)Wissenseffekte auf kognitive Fähigkeiten im Kindesalter. *Forschungskolloquium – Pädagogische Psychologie*. Philosophische Fakultät. Universität Rostock, 2016.

7.4 Inhalt der beiliegenden CD

Daten

F1 Faktorielle Differenziertheit

F2 Messinvarianz

F3 Autoregressive Effekte

F4 Kreuzverzögerte Effekte

Ergänzungen I

Vorliegende Dissertationsschrift in digitaler Form